

UMA ABORDAGEM PARA IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS A PARTIR DE APLICATIVOS PARA REDES SOCIAIS

Fábio Paschoal Júnior

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Civil.

Orientador: Nelson Francisco Favilla Ebecken

Rio de Janeiro

Junho de 2017

UMA ABORDAGEM PARA IDENTIFICAÇÃO DE PADRÕES
COMPORTAMENTAIS A PARTIR DE APLICATIVOS PARA REDES SOCIAIS

Fábio Paschoal Júnior

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA CIVIL.

Examinada por:

Prof. Nelson Francisco Favilla Ebecken, D.Sc.

Prof^a. Solange Guimarães, D.Sc.

Prof. Eduardo Bezerra da Silva, D.Sc.

Prof. Daniel Cardoso Moraes de Oliveira, D.Sc.

Prof. Eduardo Soares Ogasawara, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

JUNHO DE 2017

Paschoal Júnior, Fábio

Uma Abordagem para Identificação de Padrões Comportamentais a partir de Aplicativos para Redes Sociais / Fábio Paschoal Júnior. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIV, 111 p.: il.; 29,7 cm.

Orientador: Nelson Francisco Favilla Ebecken

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Civil, 2017.

Referências Bibliográficas: p. 101-111.

1. Redes Sociais. 2. Classificação. 3. Aplicativos Sociais. I. Ebecken, Nelson Francisco Favilla. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

*A minha mãe Waldires (In memoriam),
a minha família e aos meus amigos,
por tudo o que representam em minha vida.*

À Deus, por tornar tudo possível;

À minha família, por ser o meu pilar mestre;

Ao meu orientador, Professor Nelson Ebecken, por sua valiosa orientação, confiança, amizade e exemplo na minha vida acadêmica;

Ao meu amigo Renato Mauro, pela colaboração com a tese;

Aos meus amigos Gabriel Ribeiro, Leandro Daquer e João Antonio Ferreira, com as inúmeras sugestões e contribuições à tese;

Aos meus amigos, por todo o amor, companheirismo, confiança e paciência dedicados a mim incondicionalmente;

Aos membros da banca examinadora com suas valiosas contribuições de melhoria da tese;

AO CNPq, pela concessão da bolsa de doutorado entre 2012 e 2016, que me permitiu realizar a tese,

Agradeço.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

UMA ABORDAGEM PARA IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS A PARTIR DE APLICATIVOS PARA REDES SOCIAIS

Fábio Paschoal Júnior

Junho/2017

Orientador: Nelson Francisco Favila Ebecken

Programa: Engenharia Civil

AS Redes Sociais Online disponibilizam informações valiosas que possibilitam o estudo do comportamento de seus usuários. Algumas dessas informações são produzidas por Aplicativos Sociais, relatando informações complementares e relevantes com propriedades associadas aos usuários. Dentre as Redes Sociais Online, o Facebook é uma das mais utilizadas e se destaca pela quantidade de usuários, além permitir a associação dos Aplicativos Sociais aos perfis de seus usuários. Essa quantidade expressiva de usuários produz muitos conteúdos e são necessárias formas eficientes de coleta e de análise desses conteúdos. Dessa forma, essa tese propõe um *Framework* de coleta de dados e de análise, para classificar os usuários e estabelecer os Padrões Comportamentais dos Usuários, a partir dos conteúdos produzidos pelos Aplicativos Sociais utilizados pelos usuários. As análises são beneficiadas por um *Workflow* de execução paralela e distribuída, possibilitando a sua execução em um tempo reduzido e com escalabilidade. Assim, os usuários foram classificados conforme a sua prática de atividades físicas. Essa classificação ocorreu com a coleta e análise dos compartilhamentos de atividades físicas feitos nos perfis dos usuários, por meio do uso de Aplicativos Sociais de monitoramento de atividades físicas. Com isso, foi feita a definição do Padrão Comportamental do Usuário, relativo ao seu comportamento saudável de execução de atividades físicas.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

AN APPROACH FOR IDENTIFICATION OF BEHAVIORAL PATTERNS FROM SOCIAL NETWORKS APPS

Fábio Paschoal Júnior

June/2017

Advisor: Nelson Francisco Favilla Ebecken

Department: Civil Engineering

Online Social Networks provide valuable information that allows for the study of user's behaviors. Some of this information is produced by Social Apps, reporting companion and relevant information with properties associated to users. Among Online Social Networks, Facebook is one of the most used and stands out by the number of users. Besides it allows for the association of Social Apps to the profiles of its users. This expressive number of users produce a lot of contents. Thus, efficient ways for data collection and analysis are needed. This thesis proposes a Framework for data collection and analysis. It classifies and establishes Behavioral Patterns of Users from the content produced by the Social Apps used by users. Analyzes were modeled by a parallel and distributed execution Workflow. This enables its execution in a reduced time and with scalability. Thus, the users were classified according to their practice of physical activities. This classification occurred with the collect and analysis of shared physical activity made in users' profiles, through the use of Social Apps to monitor physical activities. With this, the definition of the Behavioral Pattern of the User was made, regarding their healthy behavior of physical activities execution.

Índice

Capítulo 1 - Introdução	1
1.1 Motivação e contexto	1
1.2 Definição do problema.....	4
1.3 Hipótese	4
1.4 Objetivos	4
1.5 Contribuições	5
1.6 Organização do texto	5
Capítulo 2 - Revisão bibliográfica	7
2.1 Redes Sociais <i>Online</i>	7
2.1.1 Conceitos	7
2.1.2 Elementos	8
2.2 Monitoramento de Atividades Físicas	8
2.3 Dispositivos móveis inteligentes	9
2.3.1 Ubiquidade	10
2.3.2 Tipos de aplicativos	11
2.3.3 Bio-sensoriamento móvel.....	12
2.4 Facebook.....	16
2.4.1 Características.....	16
2.4.2 Perfis de usuários.....	17
2.4.3 Compartilhamento e expressão de emoções.....	18
2.4.4 Privacidade dos dados	18
2.5 Aplicativos Sociais	19
2.5.1 Características.....	20
2.5.2 Compartilhamento no Facebook.....	21
2.6 Coleta de dados	22
2.6.1 Coleta distribuída.....	22
2.6.2 Uso de <i>Application Programming Interface</i> (API)	23
2.6.3 APIs para Facebook.....	23
2.6.4 APIs externas ao Facebook.....	25
2.6.5 Desenvolvimento de ferramenta.....	25
2.6.6 Aplicativos para Facebook	26

2.7	Pré-processamento dos dados coletados	26
2.8	Classificação dos dados da Rede Social Facebook.....	27
2.8.1	Classificação de dados de padrões comportamentais	27
2.9	Processamento Paralelo e Distribuído	27
2.9.1	<i>Cluster</i> Computacional	28
2.9.2	<i>MapReduce</i>	28
2.9.3	Apache Spark.....	30
2.9.4	<i>Resilient Distributed Datasets</i> (RDDs)	32
2.9.5	Apache Spark <i>DataSet/DataFrame</i>	32
2.9.6	Apache Spark SQL	33
2.9.7	Álgebra de <i>Workflows</i> Científicos.....	33
Capítulo 3 - Metodologia		35
3.1	Arquitetura do <i>Framework</i>	37
3.2	Aplicação no cenário de <i>fitness</i>	38
3.3	Sedentarismo x Atividades Físicas	43
3.4	Atividades físicas como prevenção de doenças	44
3.5	Modelo de classificação de comportamento	44
Capítulo 4 - Avaliação Experimental.....		46
4.1	Modelagem dos <i>Workflows</i>	46
4.1.1	Ambiente computacional da avaliação experimental	47
4.1.2	População e Grupos	48
4.1.3	Análise do <i>Workflow</i> de Coleta dos Dados	51
4.1.4	<i>Workflow</i> de Análise Comportamental.....	56
4.2	Resultados das Análises Comportamentais com a identificação do PCU	67
4.3	Avaliação de desempenho no processamento paralelo e distribuído.....	73
4.3.1	Volume de Dados	73
4.3.2	Análise de resultados	75
4.4	Discussões.....	80
Capítulo 5 - Trabalhos Relacionados.....		82
5.1	Análise comportamental em Redes Sociais	82
5.1.1	Predição de status de saúde com o uso de Apache Spark com <i>Machine Learning</i>	82
5.1.2	Influência das mídias sociais nos comportamentos saudáveis de jovens adultos	83

5.1.3	Mídias sociais <i>online</i> para aumentar o nível de atividades físicas	83
5.1.4	Intervenção de atividades físicas baseadas em mídias sociais	84
5.1.5	Análise de conjuntos de dados de redes sociais	84
5.1.6	Bem estar em redes sociais.....	85
5.2	<i>Framework</i> para análise de comportamento em Redes Sociais.....	86
5.2.1	Usuários Novatos x Veteranos de AS de monitoramento de atividades físicas	86
5.2.2	Motivação de atividades físicas no trabalho.....	87
5.3	Ferramentas de análise de <i>fitness</i>	88
5.3.1	MovaMais.....	88
5.3.2	Escolha de amigo para a prática de atividade física	89
5.3.3	Saúde baseada na Internet.....	90
5.3.4	Identificação e compreensão das áreas desportivas urbanas	90
5.4	Aplicativos Sociais de monitoramento de atividades físicas	91
5.5	Considerações Finais	91
Capítulo 6 - Conclusões		93
6.1	Contribuições e Publicações	94
6.2	Trabalhos Futuros	96
6.2.1	Visualizações de classificação dos usuários.....	96
6.2.2	Mineração dos Dados de Geo-referenciamento.....	98
6.2.3	Predição do Comportamento do Usuário.....	98
6.2.4	Evolução temporal de histórico do Usuário	98
6.2.5	Uso de <i>Machine Learning</i>	99
6.2.6	Coleta de dados complementares com Jsoup	99
Referências bibliográficas		101

Lista de Figuras

Figura 1 – Bio-sensor de GPS para <i>smartphones</i> (Fedkiw 2012)	12
Figura 2 – Bio-sensor de acelerômetro para <i>smartphones</i> (Fedkiw 2012).....	13
Figura 3 – Detecção dos movimentos pelo bio-sensor de acelerômetro para <i>smartphones</i> (Schirmer e Höpfner 2011).....	13
Figura 4 – Bio-sensor de giroscópio para <i>smartphones</i> (Fedkiw 2012).....	14
Figura 5 – Detecção dos movimentos pelo bio-sensor de acelerômetro em conjunto com giroscópio para <i>smartphones</i> (Schirmer e Höpfner 2011).....	14
Figura 6 – Detecção da frequência cardíaca obtida pelo aplicativo para <i>smartphones</i> ..	14
Figura 7 – Bio-sensor de monitor cardíaco <i>bluetooth</i> para <i>smartphone</i>	15
Figura 8 – Bio-sensor de monitor cardíaco para <i>smartphone</i>	15
Figura 9 – Categoria simplificada de aplicativos sociais do Facebook: Saúde e Fitness reclassificado como uma subcategoria de Jogos.	21
Figura 10 – Coleta distribuída	23
Figura 11 – Arquitetura dos Aplicativos Sociais do Facebook	26
Figura 12 – Funcionamento do <i>MapReduce</i>	29
Figura 13 – Hierarquia do Apache Spark	31
Figura 14 – Processos de Análises	35
Figura 15 – Diagrama de Classes do <i>Framework</i>	36
Figura 16 – Arquitetura do <i>Framework</i>	38
Figura 17 – Esquema de funcionamento do Arcabouço de Coleta dos Dados.....	41
Figura 18 – Postagens de AS de monitoramento de atividades físicas	42
Figura 19 – <i>Rankings</i> de usuário do arcabouço com a sua postagem no Facebook	43
Figura 20 – <i>Cluster</i> da EIC.....	48
Figura 21 – Modalidades de <i>Rankings</i> gerados	52
Figura 22 – Modalidades de <i>Rankings</i> e Grupos de Usuários	52
Figura 23 – Usuários com <i>Rankings</i> Válidos e Faixa Etária	53
Figura 24 – Regiões Geográficas e Escolaridade dos Usuários	53
Figura 25 – Estado Civil e Pessoas no <i>Ranking</i> dos Usuários.....	54
Figura 26 – Configurações e Quantidade de <i>Ranking</i> dos Usuários	54
Figura 27 – Modalidades de <i>Ranking</i> e Histórico de Atividades dos Usuários	55
Figura 28 – Tempo de utilização do FitRank e Total de Atividades dos Usuários	56

Figura 29 – <i>Schema</i> e quantidade de tuplas de DF_post_fitness.....	58
Figura 30 – <i>Workflow</i> da metodologia	59
Figura 31 – <i>Schema</i> e quantidade de tuplas de DF_Intensidades.....	60
Figura 32 – <i>Schema</i> e quantidade de tuplas de DF_Intensidades_Semanais	61
Figura 33 – <i>Schema</i> e quantidade de tuplas de DF_PCUs_Semanais	62
Figura 34 – <i>Schema</i> e quantidade de tuplas de DF_Tipos_PCUs_Semanas	63
Figura 35 – <i>Schema</i> e quantidade de tuplas de DF_PCUs	63
Figura 36 – <i>Schema</i> e quantidade de tuplas de DF_PCUs_Resumos.....	64
Figura 37 – <i>Schema</i> e quantidade de tuplas de DF_PCUs_Generos	65
Figura 38 – <i>Schema</i> e quantidade de tuplas de DF_Estatisticas_Generos	66
Figura 39 – <i>Schema</i> e quantidade de tuplas de DF_Totalizacoes.....	66
Figura 40 – População de usuários e PCU	67
Figura 41 – PCU por gêneros e Semanas Ativas dos Usuários	68
Figura 42 – PCUs das Semanas Ativas dos Usuários.....	68
Figura 43 – Classes Esportivas dos Usuários	69
Figura 44 – Atividades Físicas dos Usuários	70
Figura 45 – Intensidade e Modalidade das Atividades Físicas.....	70
Figura 46 – Modalidades e Velocidades Médias.....	71
Figura 47 – Distâncias Percorridas e Durações	71
Figura 48 – <i>Speedup</i> de Execução	78
Figura 49 – Eficiência de execução.....	79
Figura 50 – Tempo de execução.....	79
Figura 51 – Possíveis visualizações.....	97

Lista de Tabelas

Tabela 1 – Recomendações da WHO sobre a prática de atividades físicas	45
Tabela 2 – População e Grupos de Usuários	49
Tabela 3 – População de Usuários e Grupos Estudados.....	51
Tabela 4 – Replicação dos Dados.....	76
Tabela 5 – Médias das execuções para o <i>Speedup</i>	77
Tabela 6 - Resumo dos trabalhos relacionados	92
Tabela 7 – Publicações relacionadas com a tese	94

Lista de Siglas

API	<i>Application Programming Interface</i>
AS	Aplicativos Sociais
BI	<i>Business Intelligence</i>
CEFET/RJ	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
CSV	<i>Comma-Separated Values</i>
EIC	Escola de Informática & Computação
FQL	<i>Facebook Query Language</i>
GB	Giga Bytes
GHz	Giga Hertz
GPS	<i>Global Positioning System</i>
HQL	<i>Hive Query Language</i>
iOS	<i>iPhone Operating System</i>
JSON	<i>JavaScript Object Notation</i>
MB	Mega Bytes
MS	Muito Saudável
PCU	Padrões Comportamentais dos Usuários
PHP	acrônimo recursivo para " <i>PHP: Hypertext Preprocessor</i> "
PS	Pouco Saudável
RAM	<i>Random Access Memory</i>
RDD	<i>Resilient Distributed Dataset</i>
RFID	<i>Radio-Frequency IDentification</i>
RPC	<i>Remote Procedure Call</i>
RSO	Redes Sociais <i>Online</i>
S	Saudável
SQL	<i>Structured Query Language</i>
SSH	<i>Secure Shell</i>
WHO	<i>World Health Organization</i>
XML	<i>Extensible Markup Language</i>

Capítulo 1 - Introdução

As Redes Sociais *Online* (RSO) tornaram-se uma fonte de dados de valor inestimável, contendo uma riqueza de informações disponibilizadas por seus usuários. Assim, o estudo das RSO é motivado por diversas áreas do saber, seja na área de marketing, na economia, na política, na psicologia, na sociologia, na antropologia e na computação, dentre outras áreas (Benevenuto 2010, Silva 2015).

Como a humanidade está em constante evolução, os seus indivíduos criam e usam novas tecnologias que possibilitam simplificar e otimizar as tarefas de seu dia a dia. Dessa forma, há a possibilidade de ter uma melhor qualidade de vida, associada com experiências de vida mais ricas em conteúdos (Paschoal Jr et al. 2016a).

A comunicação e a interação entre as pessoas podem ser amplificadas com o uso das RSO, derrubando as barreiras de localidade de forma que a distância entre essas pessoas torna-se algo insignificante. Com isso, o volume de dados oriundos das RSO cresce e são necessárias técnicas computacionais eficientes para a recuperação de conteúdos disponibilizados nas RSO (Chau et al. 2007).

O crescimento desse volume de dados é auxiliado pelo uso de dispositivos móveis que trabalham em conjunto com as RSO (Péres et al. 2015) e com o biosensoriamento. Esse uso em conjunto possibilita o desenvolvimento de aplicativos inovadores para serem utilizados para diferentes fins, os quais disponibilizam informações relevantes de seus usuários em relação ao uso desses aplicativos (Paschoal Jr et al. 2016b).

Dentre esses aplicativos inovadores temos os aplicativos que trabalham em conjunto com as RSO, os quais possibilitam um maior nível de interação do usuário nessas RSO. Esses aplicativos são considerados como Aplicativos Sociais (AS) por esta tese, onde os mesmos relatam e disponibilizam as propriedades de seus usuários, em seus perfis nas RSO, sobre as atividades cotidianas executadas pelos mesmos (Wang et al. 2011).

1.1 Motivação e contexto

O uso crescente das RSO, seja no aumento da quantidade de usuários que elas possuem, seja na maior quantidade de tempo que esses usuários as utilizam, possibilita diversos

tipos de análises do comportamento humano (Alves e Antunes 2015). Seus usuários compartilham muitas informações entre si e, de certa forma, podem influenciar os comportamentos uns dos outros (Joinson 2008).

As RSO têm um grande potencial para influenciar a mudança de comportamento dos usuários (Alves e Antunes 2015, Oliveira e Painho 2015). Eles podem passar a ter comportamentos mais saudáveis, dada a exposição e motivação encontrada no ambiente dessas redes sociais (Morris et al. 2011, Olsen e Kraft 2009). Essa exposição e motivação, porém, podem influenciar os comportamentos de intolerâncias ou de exageros, como comportamentos esportivos excessivos que podem sugerir um comportamento pouco saudável.

Devido ao constante crescimento do nível de socialização nas RSO (Barbosa et al. 2013), estimulado pelo uso dos AS (Mutawa et al. 2012) e pelas interações com os registros das atividades de seus usuários, é possível identificar os Padrões Comportamentais dos Usuários (PCU) (Paschoal Jr et al. 2015a).

Dessa forma, a utilização dos AS nos *smartphones*, juntamente com as tecnologias de geo-referenciamento (*Global Positioning System* (GPS), acelerômetro e giroscópio) e de bio-sensoriamento (sensores de vestir, monitores cardíacos, relógios e pulseiras inteligentes) disponibilizam informações muito relevantes no monitoramento das atividades do usuário (Maksimovic et al. 2015, Péres et al. 2015) para permitir a definição dos PCU (Paschoal Jr et al. 2015b, Paschoal Jr e Ebecken 2014).

A definição dos PCU pode ser um estímulo para motivar a mudança comportamental do usuário, permitindo que ele passe a ter um comportamento mais saudável, reduzindo seu sedentarismo e também a possibilidade de ter doenças vinculadas ao comportamento sedentário (Paschoal Jr et al. 2016a).

Portanto, o estudo das RSO é repleto de desafios, dado o estágio atual das pesquisas na área e a grande quantidade de dados para a extração de conhecimento e para a mineração de dados (Benevenuto 2010). Além disso, há a complexidade de combinar as áreas interdisciplinares como a Engenharia de Software, Algoritmos, Inteligência Artificial, *BigData*, Recuperação e Visualização de Informações.

A popularidade das principais RSO cresce a todo instante. No caso do Facebook, a Argentina e o Brasil são alguns dos países com usuários ativos que mais usam a rede social (Barbosa et al. 2013, comScore 2012a, 2012b). Isso proporciona um vasto material de pesquisa e análise de dados e a extração de informações do comportamento dos usuários dessas redes sociais.

O nível de sociabilidade do Facebook faz com que o compartilhamento de conteúdos feito por seus usuários provoque um contágio emocional, relativo aos sentimentos despertados por esse conteúdo, em seus contatos desta rede social (Barbosa et al. 2013, Kramer 2012).

Esse contágio emocional ocorre tanto pela comunicação textual quanto pela comunicação indireta feita pelos meios de comunicação adicionais desta rede social, podendo ser de forma positiva ou de forma negativa (Kramer 2012). É de forma positiva quando os conteúdos compartilhados despertam sentimentos que motivam os usuários e os tornam mais solidários, ou até mesmo participativos, ao assunto envolvido no compartilhamento (Paschoal Jr e Ebecken 2014). É negativa quando despertam sentimentos nos usuários sobre intolerâncias de uma forma geral como, por exemplo, religiosa, racial, política, esportiva ou cultural (Paschoal Jr et al. 2016a) .

Os compartilhamentos dos AS esportivos podem provocar um contágio emocional positivo, pelo fato de poder motivar os usuários à prática esportiva. Isso pode ocorrer na forma de comentários, ou curtidas ou até mesmo com a prática de novas atividades físicas, por parte dos usuários contagiados positivamente. Assim, pode ser um motivador à mudança para um comportamento mais saudável.

Apesar dos AS proporcionarem esse contágio emocional positivo, eles deixam um pouco a desejar, porque não trabalham com o estudo do comportamento do usuário. Eles simplesmente relatam a execução de atividades físicas e, na melhor das hipóteses, fornecem um *ranking* de classificação de corridas com os amigos que utilizam o mesmo AS.

Nesse sentido, há a carência de um AS que permita que o usuário acompanhe o seu comportamento saudável, com a consideração de mais de uma modalidade de atividade física, não se limitando apenas as corridas. E que o usuário tenha uma análise com mais detalhes, indo além de um *ranking* de posições de corridas.

É importante a definição e o acompanhamento do comportamento do usuário. O usuário pode acreditar que está tendo um ganho de saúde, com a sua atual prática de atividades físicas quando, na verdade, pode estar tendo um comportamento pouco saudável apoiado por algum tipo de excesso.

1.2 Definição do problema

Os usuários do Facebook compartilham muitas informações, com os mais variados conteúdos e de diversas fontes diferentes. Com base nessa premissa, esta tese se propõe a abordar os seguintes problemas:

- 1.) Como é possível classificar os usuários conforme as características de seus perfis, de forma automatizada e disponível publicamente?
- 2.) Como é possível classificar os usuários conforme a utilização de AS, levando em consideração o conteúdo criado a partir desses aplicativos, de forma automatizada e disponível publicamente?
- 3.) Como é possível fazer uma análise que envolva a evolução temporal de uso dos AS para estabelecer os PCUs, de forma automatizada e disponível publicamente?

1.3 Hipótese

Esta tese busca avaliar a seguinte hipótese:

A evolução do uso dos aplicativos sociais no Facebook pode identificar padrões comportamentais dos usuários no contexto de fitness.

A identificação dos PCUs, no contexto de fitness, é importante para possibilitar um comportamento mais saudável com uma prática consciente de atividades físicas, de forma a motivar o combate ao sedentarismo com o apoio das RSO.

1.4 Objetivos

O objetivo principal desta tese é propor uma arquitetura de coleta dos dados, com a escalabilidade na classificação de dados do Facebook, na forma de um *Framework*, para classificar os seus usuários conforme perfis de utilização de AS com interação com o Facebook.

Com isso, esta tese busca:

- usar a coleta de dados para avaliar o padrão de evolução dos AS que os usuários do Facebook estão utilizando;
- classificar os perfis de usuários;
- fazer a análise comportamental considerando a evolução temporal do uso do AS, de forma a disponibilizar essas informações.

1.5 Contribuições

Esta tese espera contribuir com a disponibilização de um *Framework* capaz de oferecer mecanismos para classificações dos usuários do Facebook, por meio do uso de seus Aplicativos Sociais. Para isso, um AS de Coleta dos Dados foi disponibilizado publicamente para permitir uma competição social saudável entre seus usuários. Essa competição social pode motivar a prática de atividades físicas, de forma agradável, com o apoio das RSO e integrando as informações oriundas dos compartilhamentos de diferentes AS de monitoramento de atividades físicas.

A Coleta dos Dados ocorre somente com a utilização do AS de Coleta de Dados pelos usuários e por seus amigos, de forma a motivar a competição social entre os mesmos. Assim, com o acesso de um usuário é feita a coleta dos dados de suas atividades físicas e também é feita a coleta dos dados das atividades físicas de seus amigos.

Nesse sentido, esse *Framework* espera disponibilizar suas análises com a escalabilidade necessária ao volume de dados que for demandado, usando as técnicas atuais de processamento paralelo e distribuído.

A contribuição principal será de permitir a análise do PCU, de forma paralela e distribuída, com escalabilidade e com o uso de *Workflow* analítico. As contribuições secundárias serão oferecer uma coleta de dados com a motivação de uso da competição social saudável, bem como ser uma ferramenta de apoio ao combate ao sedentarismo.

1.6 Organização do texto

Esta tese está organizada em 6 capítulos. Este primeiro capítulo apresentou a introdução da tese, indicando a motivação, o contexto, a definição do problema, a hipótese, os objetivos e as contribuições a serem alcançadas no desenvolvimento da tese.

O Capítulo 2 apresenta a revisão bibliográfica, destacando os conceitos, as tecnologias e técnicas relacionadas a esta tese. O Capítulo 3 apresenta o *Framework* desenvolvido por esta tese, incluindo o seu processo de análises, sua arquitetura, a sua aplicação no cenário de *fitness*, os benefícios que serão associados ao usuário e os modelos de classificações utilizados na definição do PCU do usuário.

O Capítulo 4 apresenta a avaliação experimental desta tese, incluindo, o *cluster* computacional utilizado, a instanciação do *Framework* e a análise de desempenho do *Workflow* de execução, em conjunto com todas as discussões dos resultados obtidos pelas análises.

O Capítulo 5 apresenta os trabalhos relacionados e publicados que permitem a comparação do estudo desta tese, com os mesmos agrupados em três categorias com finalidades complementares. O Capítulo 6 finaliza a tese e apresenta as considerações finais, com suas conclusões, contribuições e trabalhos publicados, bem como as sugestões de trabalhos futuros.

Capítulo 2 - Revisão bibliográfica

Este capítulo descreve as principais tecnologias e conceitos computacionais que deram base para o desenvolvimento desta tese.

2.1 Redes Sociais *Online*

As RSO podem ser definidas de uma forma mais restrita, contextualizada como um serviço da Internet onde os seus usuários constroem perfis pessoais, de forma pública ou parcialmente pública. Eles interagem com outros usuários dessa rede social, compartilhando informações e conexões com o seus contatos (Benevenuto 2010, boyd e Ellison 2008, Mislove 2009).

Apesar das RSO terem alcance mundial, elas são influenciadas com as questões de diferenças etnográficas de comportamento de seus usuários. Conforme a localidade, mudam as formas de interação e de publicação de conteúdos, bem como os objetivos típicos de utilização das RSO e a forma de se auto expressar (Chapman e Lahav 2008).

Atualmente, temos por volta de 210 RSO com os mais variados propósitos (Wikipedia 2017). Dentre elas, uma das mais utilizadas é o Facebook (Cosenza 2017) e, por esse motivo, é o objeto de estudo desta tese. Portanto, as publicações dos AS das atividades físicas compartilhadas no Facebook contêm os dados que permitiram a definição dos PCU propostos por esta tese (Paschoal Jr et al. 2016b).

2.1.1 Conceitos

A estrutura de uma rede social é guiada pela teoria dos grafos. Dessa forma, cada item da rede social (característica do objeto de estudo da rede social) pode ser considerado como vértice. As conexões entre esses vértices (relacionamentos entre dois vértices) são denominadas arestas (Benevenuto 2010), as quais podem representar o compartilhamento de alguma informação entre os vértices conectados pela aresta.

Em complementação, as RSO seguem também a teoria de redes complexas, devido ao seu grande número de vértices e arestas (Benevenuto 2010). E, nesse ponto, há o desafio de trabalhar o processamento de uma grande quantidade de dados, pois algumas vezes é necessária a coleta do grafo inteiro da rede social em estudo, mesmo

que limitado a alguma questão regional. Dessa forma, esse grafo pode ter milhões de vértices e bilhões de arestas (Benevenuto e Almeida 2011).

Uma forma de contornar esse problema é que pode ser construído um modelo de grafos incorporando as características pesquisadas na rede social. Nesse caso, é possível construir um sub-grafo da rede social para fazer uma predição de comportamento da rede social analisada, diminuindo sensivelmente o tamanho do grafo analisado (Cervi 2008).

A maioria das RSO é dinâmica, ou seja, o seu grafo representativo muda conforme os vértices ou arestas são adicionados ou excluídos da rede social. Com base nisso, a tendência da rede social é ficar mais densa com o passar do tempo, com o crescimento caracterizado pela lei de potência (Cervi 2008, Han et al. 2011).

2.1.2 Elementos

As RSO são formadas por alguns elementos estruturais. Os principais elementos são: perfis de usuários, atualizações de conteúdos e comentários em conteúdos. Os perfis dos usuários trazem a descrição das informações dos usuários, como dados pessoais, interesses, fotos, lista de amigos, dentre outras. As atualizações de conteúdo são formadas pela disponibilização de novos conteúdos compartilhados pelos usuários, motivando a interação entre os mesmos. Os comentários em conteúdos permitem que os usuários expressem as suas ideias, opiniões e sentimentos, tanto nos perfis dos usuários quanto nos conteúdos compartilhados (Benevenuto 2010).

2.2 Monitoramento de Atividades Físicas

Os indivíduos praticantes de atividades físicas podem ter uma vida mais saudável, devido à importância dessas atividades para o bom funcionamento do organismo. Por conta disso, muitos praticantes têm algum tipo de controle ou fazem o monitoramento da execução de suas atividades físicas. Esse monitoramento ocorre de formas distintas, como o registro de suas atividades físicas e o acompanhamento da evolução temporal de execução das mesmas (Paschoal Jr et al. 2016a).

O registro da atividade física pode conter informações detalhadas de execução da atividade e que foram obtidas através do uso de algum bio-sensor. No contexto desta

tese, o bio-sensor é um dispositivo mecânico ou eletrônico que disponibiliza algumas propriedades do indivíduo ou do meio onde a atividade física foi executada (Paschoal Jr et al. 2016a).

É possível classificar os bio-sensores em três grupos, os quais são: i.) sensores embarcados em dispositivos móveis; ii.) sensores independentes que podem ser acoplados aos dispositivos móveis e iii.) sensores independentes que não são acoplados aos dispositivos móveis, porém disponibilizam informações complementares (Paschoal Jr et al. 2016a).

Os bio-sensores considerados por essa tese são: i.) *Global Positioning System* (GPS); ii.) pedômetro; iii.) acelerômetro; iv.) giroscópio; v.) monitor cardíaco; vi.) pulseiras inteligentes; vii.) relógios inteligentes e viii.) termômetro (Paschoal Jr e Ebecken 2014).

No monitoramento de atividades físicas são obtidas as principais informações sobre a execução da atividade física. Isso ocorre através do uso de bio-sensores, os quais podem disponibilizar informações como a geolocalização, a velocidade, a distância percorrida, o ritmo cardíaco, a duração, a altitude, o declive, o apace, o número de passos, o clima e a temperatura (Paschoal Jr et al. 2016a).

Há duas formas de execução do monitoramento das atividades físicas: a forma manual e; a forma automatizada. Na forma manual é feito o registro das informações básicas, sobre a atividade física, em tabela ou planilha, obtidas por bio-sensores usados de forma não integrada, sendo de responsabilidade do usuário utilizar o bio-sensor e coletar as informações do monitoramento. Na forma automatizada é feito o registro das informações, sobre a atividade física, com o uso de dispositivos móveis inteligentes, obtidas automaticamente com o uso de bio-sensores e de seus respectivos softwares de controle, centralizando as informações coletadas em softwares específicos de monitoramento de atividades físicas, os quais são considerados nesta tese como AS de monitoramento de atividades físicas (Paschoal Jr et al. 2016a).

2.3 Dispositivos móveis inteligentes

Os dispositivos móveis inteligentes, especialmente os *smartphones* utilizando AS com as RSO, vêm tornando-se ótimas ferramentas de persuasão para estimular os seus

usuários às mudanças de estilo de vida, para passar a ter um estilo de vida mais saudável (Morris et al. 2011).

Essa persuasão vem sendo feita antes mesmo do uso intensivo de *smartphones*, onde os telefones móveis tradicionais foram utilizados para a motivação de grupos de amigos para que tivessem mudanças de comportamento mais saudáveis (Consolvo et al. 2006, Morris et al. 2011).

Atualmente, essa persuasão é estimulada pelas informações que são disponibilizadas pelos bio-sensores embutidos nos *smartphones*, como o GPS, o giroscópio, o pedômetro e o acelerômetro. Esse estímulo também ocorre com as informações disponibilizadas pelos bio-sensores externos conectados ao *smartphone* por protocolos de comunicação de *WiFi*, *Bluetooth* ou radiofrequência, como monitores cardíacos, pulseiras inteligentes e relógios inteligentes (Paschoal Jr et al. 2015a).

O funcionamento dos bio-sensores nos *smartphones* é coordenado por aplicativos específicos dos bio-sensores. Porém, os AS de monitoramento de atividades físicas utilizam esses bio-sensores no monitoramento da atividade física, de forma automatizada, disponibilizando em tempo real para o seu usuário informações precisas de execução da atividade física. Isso permite uma experiência de atividade física mais segura, agradável e motivadora, tanto em relação ao desempenho do usuário quanto aos riscos de saúde relacionados ao excesso de esforço físico (Paschoal Jr et al. 2016a).

2.3.1 Ubiquidade

A ubiquidade é a propriedade de estar presente a todo instante e em qualquer lugar, de forma a ser praticamente imperceptível. Assim, os *smartphones* possibilitam o pleno exercício da ubiquidade, ou seja, aplicam os conceitos de computação ubíqua, de oferecerem aplicativos para diversas atividades. Dessa forma, as barreiras computacionais são transparentes para os usuários, as tarefas complexas são feitas de forma muito simples pela interação do usuário com o aplicativo, sem que o usuário perceba o sistema computacional e toda a complexidade embutida no mesmo (Domingues 2008, Péres et al. 2015).

A computação ubíqua é formada pela interseção entre a computação móvel e a computação pervasiva. Isso significa que ela integra a mobilidade com sistemas

distribuídos, inteligentes, imperceptíveis e com integração com outros computadores (Domingues 2008).

2.3.2 Tipos de aplicativos

Atualmente, há uma grande variedade de aplicativos disponíveis para *smartphones*. Eles podem ser obtidos nas lojas on-line de aplicativos, de acordo com a plataforma utilizada pelo smartphone (Android¹, iOS ou Windows Phone).

Alguns aplicativos são gratuitos, normalmente com recursos mais simples, mas há também os aplicativos pagos, a um preço bem acessível, com versões mais elaboradas e com a disponibilidade de mais recursos do que os aplicativos gratuitos.

Para esta tese são considerados os aplicativos que tenham interação com o Facebook, especificamente os AS de monitoramento de atividades físicas, os quais permitem o compartilhamento das informações do uso do aplicativo no perfil do usuário, com uma grande riqueza de informações.

Segundo o Google Play Store (Google 2017), alguns dos AS de monitoramento de atividades físicas com uma maior quantidade de usuários são o Endomondo Corrida Ciclismo MTB (Endomondo 2017a, 2017b), MapMyFitness (MapMyFitness 2017a, 2017b), MapMyRide Equitação Ciclismo (MapMyRide 2017a, 2017b), MapMyRun (MapMyRun 2017a, 2017b), MapMyWalk (MapMyWalk 2017a, 2017b), Nike+ Running (Nike 2017a, 2017b), Polar Beat Fitness Coach (Polar Beat 2017a, 2017b), Polar Flow Activity & Sports (Polar Flow 2017a, 2017b), RunKeeper GPS Correr e Caminhar (RunKeeper 2017a, 2017b), Runtastic Corrida e Caminhada (Runtastic 2017a, 2017b), Runtastic PRO GPS Correr (Runtastic 2017c), Runtastic Road Bike Ciclismo (Runtastic 2017d), Runtastic Road Bike PRO GPS (Runtastic 2017e), Runtastic Mountain Bike GPS (Runtastic 2017f), Runtastic Mountain Bike PRO (Runtastic 2017g) e Strava GPS Correr Ciclismo (Strava 2017a, 2017b).

¹ <https://play.google.com/store>

2.3.3 Bio-sensoriamento móvel

Os *smartphones* possibilitam tanto a conexão com a Internet quanto com outros sistemas computacionais. Muitos aplicativos para *smartphones* acessam informações complementares disponibilizadas em sítios da Internet e em sistemas computacionais específicos de geo-referenciamento, com o uso da comunicação de dados móveis desses dispositivos.

A maior parte dos *smartphones* já possui sensores especiais embutidos ou sensores especiais que possam ser acoplados ao aparelho. Alguns desses sensores podem medir propriedades do usuário e, nesta tese, serão considerados como bio-sensores.

Com base nisso, o bio-sensoriamento móvel consiste na disponibilização de propriedades do usuário. Essas propriedades relatam características temporais do usuário, com a utilização de aplicativos móveis específicos, que enriquecem a experiência de uso do aplicativo e detalham mais os dados relativos à atividade realizada pelo usuário.

No contexto desta tese, os principais bio-sensores são GPS, o pedômetro, o acelerômetro, o giroscópio, o monitor cardíaco e o *Radio-Frequency IDentification* (RFID).

O GPS é um sensor embutido no *smartphone* que permite obter a localização geográfica do usuário, através do geo-referenciamento, incluindo o deslocamento efetuado pelo mesmo (Pantas 2013). O sensor de GPS é passivo, ou seja, somente recebe as informações do sistema de GPS e não transmite informações para o satélite, apenas emite um sinal periódico indicando o horário e a posição atual. A Figura 1 ilustra o sensor de GPS que é embutido nos *smartphones* (Fedkiw 2012).



Figura 1 – Bio-sensor de GPS para *smartphones* (Fedkiw 2012)

O pedômetro é posicionado na cintura do usuário, através de um cinto, o qual detecta as passadas (pisadas) que o usuário faz durante o seu deslocamento. É um

dispositivo de baixo custo e possibilita medir a distância percorrida pelo usuário (Hurling et al. 2007).

O acelerômetro tem sido utilizado amplamente em monitoramento de atividades físicas, independentemente da faixa etária do usuário (Hurling et al. 2007). É um sensor muito mais preciso e sensível que o pedômetro e, por isso, está embutido nos *smartphones*. Ele faz o monitoramento da mudança de direção e/ou ângulo do *smartphone*, incluindo a sua velocidade de movimento (Schirmer e Höpfner 2011). Trabalha no espaço tridimensional e é muito utilizado pelos aplicativos do *smartphone*, como orientação da tela (retrato ou paisagem), jogos com controle de movimentos (guiar um veículo) ou aplicativos que detectam o movimento do usuário (Pantas 2013). A Figura 2 ilustra o bio-sensor de acelerômetro (Fedkiw 2012).



Figura 2 – Bio-sensor de acelerômetro para *smartphones* (Fedkiw 2012)

A Figura 3 ilustra os movimentos do *smartphone* que são detectados pelo bio-sensor de acelerômetro (Schirmer e Höpfner 2011).

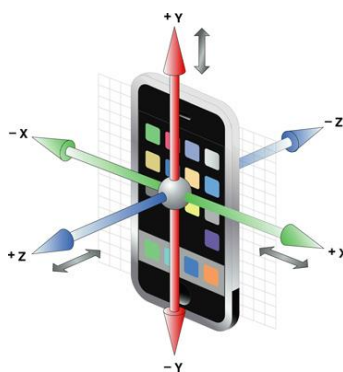


Figura 3 – Detecção dos movimentos pelo bio-sensor de acelerômetro para *smartphones* (Schirmer e Höpfner 2011)

O giroscópio é um sensor embutido nos *smartphones*. Ele faz o monitoramento da mudança de direção e/ou posição do *smartphone*. A utilização em conjunto com o acelerômetro permite a detecção de 6 movimentos diferentes (Pantas 2013, Schirmer e Höpfner 2011). A Figura 4 ilustra o bio-sensor de giroscópio (Fedkiw 2012).



Figura 4 – Bio-sensor de giroscópio para *smartphones* (Fedkiw 2012)

A Figura 5 ilustra os movimentos do *smartphone* que são detectados pelo bio-sensor de acelerômetro em conjunto com giroscópio (Schirmer e Höpfner 2011).

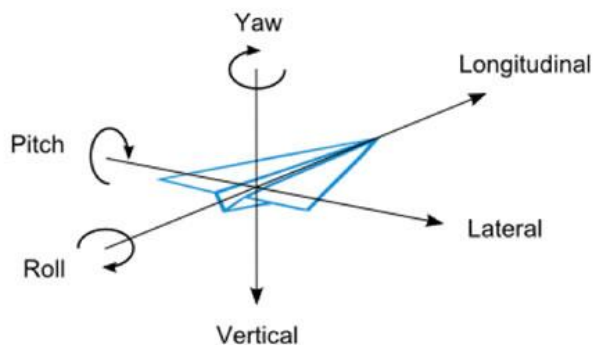


Figura 5 – Detecção dos movimentos pelo bio-sensor de acelerômetro em conjunto com giroscópio para *smartphones* (Schirmer e Höpfner 2011)

O monitor cardíaco é um sensor que começa a ser embutido nos *smartphones*. Ele faz o monitoramento da frequência cardíaca do usuário. As informações de monitoramento podem ser postadas nas redes sociais, caso seja de interesse do usuário. Há modelos para serem utilizados durante a atividade física e modelos a serem utilizados fora da atividade física.

A Figura 6 ilustra a publicação no Facebook do resultado do monitoramento cardíaco obtido pelo aplicativo *Runstatic Heart Rate*.



Figura 6 – Detecção da frequência cardíaca obtida pelo aplicativo para *smartphones*

A Figura 7 ilustra o bio-sensor de monitor cardíaco para ser utilizado em um *smartphone* pela conexão *bluetooth*. Ele deve ser fixado no tórax e é utilizado para monitorar a frequência cardíaca durante a execução da atividade física, com a sua

medição feita através de alguns aplicativos da família de aplicativos *Runtastic PRO*² (Runtastic Shop 2017).



Figura 7 – Bio-sensor de monitor cardíaco *bluetooth* para *smartphone*

A Figura 8 ilustra o bio-sensor de monitor cardíaco para ser utilizado em um *smartphone* pela conexão de fone de ouvido. O emissor deve ser fixado no tórax e o receptor deve ser fixado na conexão de fone de ouvido. Assim, é utilizado para monitorar a frequência cardíaca durante a execução da atividade física, com a sua medição feita através de alguns aplicativos da família de aplicativos *Runtastic PRO*² (Runtastic Shop 2016).



Figura 8 – Bio-sensor de monitor cardíaco para *smartphone*

O RFID é um sensor de radio frequência, contendo chips de silício e antenas, que pode ser utilizado para a identificação automática de pessoas, animais, produtos,

² <https://www.runtastic.com/pt-br/apps>

objetos e roupas, dentre outras possibilidades. Pode ser embutido em um cartão ou etiqueta eletrônica. Ele vem sendo utilizado para criar postagens automáticas no Facebook, com o auxílio de AS, no sentido de confirmar a presença do usuário em tal localidade (Hypeness 2012, Karau et al. 2015).

Dessa forma, conclui-se que essas tecnologias de bio-sensoriamento móveis utilizadas pelos AS permitem a definição de uma evolução temporal detalhada de seus usuários e, se essas informações forem publicadas nas RSO, permitem estabelecer o PCU, pois elas disponibilizam informações importantes para os mesmos de forma prática e eficaz, possibilitando que os usuários tenham uma experiência mais rica de monitoramento (Paschoal Jr e Ebecken 2014).

2.4 Facebook

O Facebook é uma das maiores redes sociais da atualidade (Cosenza 2017) e conseguiu isso por estar em constante evolução, adaptando os seus serviços para tornar-se mais atrativa tanto para os seus atuais usuários quanto para captar novos usuários, de forma que os usuários tenham uma melhor percepção em relação à utilização desta rede social (Joinson 2008, Lampe et al. 2008).

2.4.1 Características

Apesar de o Facebook ter alcance mundial e de seus usuários terem a liberdade de adicionar novos contatos para criar novas conexões, os usuários dessa rede a utilizam, preferencialmente, para criar conexões com usuários que conheceram de forma *off-line*, seja para manter relacionamentos já existentes ou para solidificar relacionamentos recentes que se iniciaram de forma *off-line* (Ellison et al. 2007, Madden et al. 2013).

Porém, com a criação do mecanismo de grupos de usuários, tornou-se muito comum os usuários adicionarem novos contatos que são amigos de amigos ou amigos pertencentes de grupos em comum, onde a afinidade de interesses e de assuntos prevalece sobre o fato de não se conhecerem previamente de forma *off-line*.

Os mecanismos de sugestões de “pessoas que talvez você conheça” e “outras pessoas que talvez você conheça”, outros novos mecanismos que estão disponíveis aos usuários do Facebook, apresentam uma relação de amigos de amigos disponibilizada

pelo Facebook, onde é estimulada e simplificada a criação de novas conexões com usuários.

Há também a proteção contra a solicitação indiscriminada de novas amizades por parte do usuário. A primeira delas é advertência ao usuário, quando ele tenta criar conexão com uma pessoa completamente desconhecida, mesmo sendo amigo de amigo ou provavelmente com uma distância de mais de dois nós (amigo do amigo do amigo), que ele deve criar conexões apenas com pessoas que ele conhece de forma *off-line*. A segunda delas é o bloqueio de novas solicitações de amizades ao usuário, enquanto ele tiver pendentes mil solicitações de amizades não respondidas.

Outras proteções do Facebook são: i.) bloquear temporariamente o usuário que faz muitas curtidas em um período de tempo muito pequeno e para muitos usuários diferentes, como curtir uma relação de comentários de usuários diferentes em uma postagem (por exemplo, mais de 200 curtidas em menos de 2 minutos, tendo um comportamento de robô ou de quem não está lendo os comentários); ii.) Não deixar o usuário re-cutucar outro usuário enquanto esse usuário não tiver respondido o cutucar enviado anteriormente; iii.) Não deixar o usuário reenviar um pedido de amizade não respondido, onde o usuário deve cancelar o pedido enviado anteriormente para poder fazer o novo pedido; iv.) Não deixar o usuário convidar novamente outro usuário para curtir alguma página pessoal feita no Facebook e que esse outro usuário não tenha curtido.

O número médio de contatos dos usuários é em torno de 250. Desses contatos, os usuários têm interações constantes com até 40 contatos. As principais interações entre os usuários são baseadas na curiosidade de ver o que os seus contatos estão compartilhando, ou de estar mais próximo deles mantendo o contato virtual, ou até mesmo reencontrando pessoas que os usuários não tinham contato há tempos. Essas interações são percebidas pelos usuários como um momento de entretenimento, descontração, de relaxamento físico e emocional de seus afazeres diários (Olsen e Kraft 2009).

2.4.2 Perfis de usuários

Os perfis de usuários do Facebook concentram as informações pessoais de seus usuários e diversos mecanismos técnicos de compartilhamento de informações. Os principais

mecanismos descrevem os atributos pessoais (nomes, familiares, relacionamentos, sentimentos, amigos, preferências, fotos, vídeos, viagens, locais que frequentaram, hipertextos para diversos outros tipos de conteúdos, atividades executadas, eventos que participaram, atualizações de status, comentários, declarações, músicas, animações, etc.) e também é possível associar aplicativos sociais aos mesmos (Barbosa et al. 2013, boyd e Ellison 2008, Ellison et al. 2007, Olsen e Kraft 2009, Paschoal Jr et al. 2016a).

Os mecanismos técnicos de compartilhamento de informações são constantemente atualizados e incrementados com novas funcionalidades para manter, motivar, ou até mesmo ampliar, o interesse dos usuários em utilizar a rede social (Ellison et al. 2007, Joinson 2008, Kraft et al. 2008, Lampe et al. 2008, Olsen e Kraft 2009).

2.4.3 Compartilhamento e expressão de emoções

A criação de conteúdos no perfil do usuário é feita pelo mesmo e por seus contatos. O compartilhamento é feito com base na descrição dos atributos pessoais dos usuários. Os AS também são utilizados para a criação e o compartilhamento de conteúdos no perfil do usuário, de acordo com o propósito específico do aplicativo.

Esses conteúdos são postados no mural de notícias do usuário e são passíveis de interação com os seus contatos, como a expressão de emoções (curtir, amar, risada, impressionado, triste ou aborrecido), comentários, marcações e compartilhamento, mecanismos esses que incrementam a sociabilidade dos usuários na rede social. O mecanismo de bate-papo complementa essa sociabilidade (Barbosa et al. 2013, Olsen e Kraft 2009).

Porém, em algumas situações o usuário se arrepende por ter compartilhado algum conteúdo no Facebook, seja por não ser um conteúdo apropriado ou que provocou interações negativas (Wang et al. 2011).

2.4.4 Privacidade dos dados

O Facebook possui controles para ajustar os critérios de privacidade que o usuário deseja ter em seu perfil. Os usuários devem ter consciência dos riscos que eles correm com o ajuste incorreto desses controles, bem como o de ter uma atitude de

autopromoção e compartilhar informações privadas, como saúde pessoal. Os motivos para os usuários compartilharem esse tipo de informação privada variam entre suporte emocional, motivacional, aconselhamento e prestação de contas (Newman et al. 2011).

Uma abordagem interessante que não é oferecida pelo Facebook, mas que tem uma grande aceitabilidade, é permitir a atribuição de ajustes de privacidade de um determinado aplicativo social para que ele compartilhe as informações apenas para um subgrupo de amigos do usuário e não para todos os usuários. A falta desse recurso provocou preocupação dos usuários do aplicativo social “3GT – *Three Good Things*” para exercícios físicos e compartilhamento de bem estar (Munson et al. 2010).

A vulnerabilidade dos usuários, em geral, é considerada baixa. Aproximadamente 94% dos usuários ajustam adequadamente os controles de privacidade de seus perfis, aproximadamente 4% tem uma vulnerabilidade média e aproximadamente 2% tem uma alta vulnerabilidade (Barbosa et al. 2013, Silva et al. 2012).

Um dos desafios quando se deseja coletar dados do Facebook é tratar as questões de privacidade dos seus usuários. Para aumentar o nível de privacidade de seus usuários, o Facebook está constantemente aprimorando os seus mecanismos de coleta de dados e, devido a isso, a coleta de dados no Facebook tem sido um desafio cada vez maior.

Os AS que necessitem fazer a coleta de dados devem solicitar, ao menos, dois tipos de permissões. O primeiro tipo deve ser feito ao Facebook, que representa as permissões de acesso aos tipos de conteúdos que serão coletados pelos AS. O segundo tipo deve ser feito aos usuários, para que os AS tenham permissões para acessarem os dados privados dos usuários (Cheng et al. 2013, Paschoal Jr et al. 2015a).

Porém, por definição padrão do Facebook, o usuário tem como opção liberar o acesso a todos os seus dados privados ou não liberar o acesso aos seus dados privados, não havendo flexibilização de selecionar de forma parcial os dados privados a serem acessados (Cheng et al. 2013). Esses dados privados estão relacionados aos conteúdos coletados pelos AS, não sendo possível flexibilizar a liberação para o acesso parcial a esses dados privados.

2.5 Aplicativos Sociais

O Facebook permite a criação de páginas internas ou externas que façam a integração com a sua plataforma. Essas páginas são conhecidas como AS para o Facebook, os

quais possibilitam uma maior socialização de seus usuários. Assim, esses aplicativos fazem o compartilhamento de informações, dentro do perfil do usuário do Facebook, com base em informações externas da mesma.

Os AS são vinculados aos perfis dos usuários e permitem que os usuários elevem os seus níveis de interação no Facebook. Os AS criam e compartilham conteúdos nos perfis dos usuários, conforme o propósito de utilização dos AS, permitindo inferir características importantes de seus usuários, incluindo os seus comportamentos (Paschoal Jr et al. 2016b). Os conteúdos criados pelos AS no Facebook podem ser passíveis de toda a socialização com curtidas, comentários, marcações e compartilhamentos (Barbosa et al. 2013, Olsen e Kraft 2009).

O desenvolvimento de um AS é feito com a utilização da *Application Programming Interface* (API) do Facebook (Facebook 2017a), a qual disponibiliza a execução de diversas ações com o perfil dos usuários do Facebook como, por exemplo, a publicação de conteúdos, histórias e fotos, dentre outras informações.

Os AS se diferenciam dos *sites* que contém *plugins* sociais do Facebook. Em boa parte dos *sites* atuais encontramos os botões de curtir e compartilhar do Facebook e, neste caso, foram utilizados os *plugins* sociais nestes botões. A partir desses botões o usuário pode curtir e/ou compartilhar, em seu perfil no Facebook, o conteúdo do *site* que está acessando (Facebook 2017b).

2.5.1 Características

Os AS são categorizados pelo Facebook de acordo com os seus objetivos e as suas finalidades, onde há uma lista de AS disponíveis para Facebook (Facebook 2014a). A Figura 9 ilustra, de forma simplificada, essa categorização focando os AS relacionados às análises desta tese. Mesmo o Facebook tendo recentemente reclassificado os AS de atividades físicas em uma subcategoria da categoria de jogos, eles continuaram classificados como AS de saúde e *fitness* (Endomondo 2017a, MapMyFitness 2017a, MapMyRide 2017a, MapMyRun 2017a, MapMyWalk 2017a, Nike 2017a, Paschoal Jr et al. 2016b, Polar Beat 2017a, Polar Flow 2017a, RunKeeper 2017a, Runtastic 2017a, Strava 2017a).

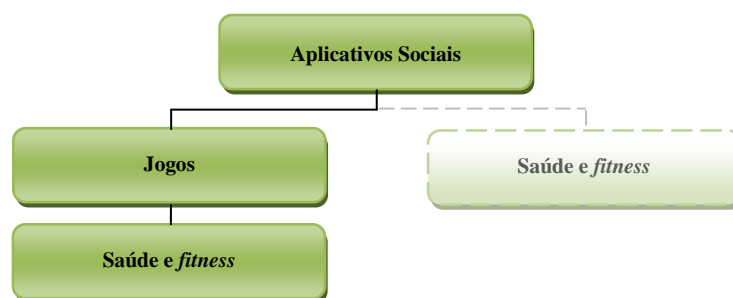


Figura 9 – Categoria simplificada de aplicativos sociais do Facebook: Saúde e Fitness reclassificado como uma subcategoria de Jogos.

Em cada categoria de aplicativos são oferecidos diversos aplicativos sociais que interagem com o Facebook, disponibilizando as informações do usuário relativas ao seu comportamento, gostos, preferências ou hábitos. Desta forma, ocorre o enriquecimento da publicação das informações dos usuários.

2.5.2 Compartilhamento no Facebook

Existem alguns formatos diferenciados de layout das informações compartilhadas nos perfis de usuários do Facebook. As diferenças se dão pela plataforma computacional utilizada no compartilhamento e/ou pelo tipo de informação que se deseja compartilhar (Facebook 2017c).

Na plataforma computacional utilizada no compartilhamento, temos as variações de quem está fazendo o compartilhamento, ou seja, se é um aplicativo para *iPhone Operating System* (iOS), ou um aplicativo para Android, ou um *site da web* ou um aplicativo para *web* (Facebook 2017c).

No tipo de informação que se deseja compartilhar, pode-se ter o compartilhamento que contenha um *link*, ou uma mensagem, ou uma atualização de *status*, ou uma estória estruturada como um histórico de uma atividade física (Facebook 2017c).

Há também o compartilhamento de estórias mais ricas e personalizadas, onde há uma estrutura de dados definida neste compartilhamento, como a pessoa que faz a publicação (ator), o que está sendo feito pela pessoa (ação), como foi feita a ação (objeto) e qual foi a aplicação que fez a postagem deste compartilhamento. Neste tipo de compartilhamento é utilizada a ferramenta *Open Graph* (Facebook 2016a).

Nas análises desta tese, temos as publicações do tipo estórias, criadas pelos AS de monitoramento de atividades físicas, as quais contêm a ação feita pelo usuário

(corrida, caminhada ou ciclismo), as informações de geo-referenciamento de execução da atividade (mapa ou rota percorrida), as informações de execução da atividade (velocidade, duração, distância percorrida, calorias, etc.) e o nome do AS de monitoramento de atividades físicas que criou a postagem (Facebook 2016a, Paschoal Jr et al. 2016a).

2.6 Coleta de dados

Essa seção apresenta as principais formas de coleta de dados das RSO que vêm sendo utilizadas para a mineração do conteúdo das RSO.

2.6.1 Coleta distribuída

A coleta distribuída de dados das RSO pode ser feita com o uso de agentes coletores distribuídos em diversos computadores. Essa forma de coleta é adequada pela grande quantidade de dados das RSO, onde as tarefas de obter e tratar os dados coletados são divididas entre esses computadores (Benevenuto 2010).

Nessa abordagem, pode ser utilizada uma máquina central (mestre) que faz o gerenciamento dos usuários que serão acessados da rede social em estudo. Em sequência, a máquina mestre divide essa relação de usuários entre máquinas auxiliares (trabalhadoras), as quais fazem a coleta dos dados dos usuários, salvam esses dados, identificam os novos usuários encontrados nos dados coletados e retornam para a máquina mestre a relação de novos usuários. Assim, a máquina mestre inicia novamente o processo de distribuição desses novos usuários para as máquinas trabalhadoras e o processo todo é repetido (Chau et al. 2007).

Com isso, evita-se a coleta redundante de dados, pois a máquina mestre faz o gerenciamento dos usuários que tiveram as informações processadas. A Figura 10 ilustra essa abordagem (Benevenuto 2010).



Figura 10 – Coleta distribuída

2.6.2 Uso de *Application Programming Interface* (API)

As APIs das RSO são compostas por um agrupamento de requisições e respostas a essas requisições. Normalmente, manipulam as informações dos usuários e de seus contatos. As principais RSO oferecem APIs para a coleta de dados e que retornam esses dados em formatos *Extensible Markup Language* (XML) ou *JavaScript Object Notation* (JSON) (Paschoal Jr e Ebecken 2014).

2.6.3 APIs para Facebook

Coletar informações das publicações dos usuários do Facebook tem se tornado um crescente desafio. Isso ocorre pelo fato do Facebook ter feito o constante aprimoramento de suas ferramentas que permitem a coleta de dados e também das constantes alterações das políticas de privacidade dos dados criados pelo Facebook. O Facebook disponibiliza suas APIs para coletar os dados, as quais possibilitam a troca de dados dos AS com o Facebook, seguindo uma forma padronizada de comunicação (Paschoal Jr e Ebecken 2014).

Essas APIs são usadas para manipular os dados de seus usuários, permitindo que os AS façam operações de escrita e de leitura nos perfis dos usuários. Porém, é necessário que os AS peçam autorizações para fazer a coleta de dados e para postar os seus conteúdos, primeiro para o Facebook, indicando o tipo de dados que será coletado/postado e o motivo, e finalmente para os usuários dos AS, onde os usuários

indicam que concordam com o uso dos dados coletados e das postagens em seus perfis (Paschoal Jr et al. 2016a). Os principais ambientes de desenvolvimento com o uso das APIs do Facebook são iOS, Android, JavaScript e acrônimo recursivo para "*PHP: Hypertext Preprocessor* (PHP) (Facebook 2017a).

As APIs do Facebook têm o seu desenvolvimento evoluído constantemente. A versão inicial, denominada v1.0, ficou em uso desde de abril de 2010 e foi suportada até abril de 2015. A versão v2.0 ficou em uso desde abril de 2014 e foi suportada até agosto de 2016. A versão v2.1 ficou em uso desde agosto de 2014 e foi suportada até outubro de 2016. A versão v2.2. ficou em uso desde outubro de 2014 e foi suportada até março de 2017. A versão v2.3 está em uso desde março de 2015 e deverá ter suporte até julho de 2017. A versão v2.4 está em uso desde julho de 2015 e deverá ter suporte até outubro de 2017. A versão v2.5 está em uso desde outubro de 2015 e deverá ter suporte até abril de 2018. A versão v2.6 está em uso desde abril de 2016 e deverá ter suporte até julho de 2018. A versão v2.7 está em uso desde julho de 2016 e deverá ter suporte até outubro de 2018. A versão v2.8 está em uso desde outubro de 2016 e deverá ter suporte até abril de 2019. Atualmente, temos a versão v2.9, a qual está em uso desde abril de 2017 e será suportada até outubro de 2019 (Facebook 2017d). A proposta do Facebook é dar suporte a cada versão de suas APIs por 2 anos (Facebook 2017e).

A principal API do Facebook é a "*Graph API*". Ela é a que é utilizada atualmente, possibilita a execução das principais ações de postagens e de recuperação de dados dos perfis dos usuários (Facebook 2017a). Existem outras APIs do Facebook, como a "*Public Feed API*", a "*Keyword Insights API*" e a "*Chat API*" (Facebook 2017f).

A "*Public Feed API*" possui, atualmente, acesso restrito a alguns parceiros do Facebook. Essa API disponibiliza um fluxo de atualizações de *status* e de páginas dos usuários desta rede social (Facebook 2017g).

Já a "*Keyword Insights API*" possui, atualmente, acesso restrito a alguns parceiros do Facebook. Essa API disponibiliza uma análise sobre o quanto uma palavra chave ou frase foi utilizada em publicações dos usuários (atualização de *status*, legenda de hipertexto, legenda de imagem e comentários postados). Essa análise ainda pode ser classificada com características dos usuários, como faixa etária ou cidade em que reside (Facebook 2014b).

A "*Chat API*" permite integrar o *chat* do Facebook com outros aplicativos de mensagens instantâneas, porém, ela está atualmente em desuso e sem suporte do

Facebook, pois foi substituída pela criação do Messenger integrado ao Facebook (Facebook 2016b).

Em complementação, há a “*Facebook Query Language (FQL)*”, a qual disponibiliza alguns comandos no estilo *Structured Query Language (SQL)* para consultar os dados da *Graph API* (Facebook 2016c). Porém, a FQL não está mais disponível desde agosto de 2016, a qual foi incorporada à *Graph API* (Facebook 2017d).

Os AS são criados com o auxílio da *Graph API*, permitindo aos mesmos a publicação e a coleta de conteúdos nos perfis dos usuários no Facebook. Isso dá a possibilidade de o AS publicar conteúdos, histórias e fotos nos perfis de seus usuários, bem como coletar os conteúdos de publicações e das propriedades dos usuários (Facebook 2017a).

2.6.4 APIs externas ao Facebook

A API RestFB não pertence ao Facebook, foi escrita em Java, trabalha de forma flexível em conjunto com a *Graph API* e facilita o trabalho de desenvolvimento por apresentar uma robustez em relação as mudanças frequentes que ocorrem nas APIs do Facebook (Allen e Bartels 2017).

A maioria dos AS de monitoramento de atividades físicas postam os seus conteúdos nos perfis dos usuários do Facebook de forma bem resumida, porém contendo um hipertexto para o *site* externo do AS, o qual detalha melhor a execução das atividades físicas. Como essas informações detalhadas não estão no Facebook, elas podem ser coletadas com o uso de outra API não pertencente ao Facebook. Então, a API Jsoup possibilita a coleta de dados de páginas da *web*, é uma biblioteca feita em Java e pode ser utilizada para coletar essas informações detalhadas disponibilizadas no hipertexto para o *site* externo do AS (Hedley 2017, Paschoal Jr et al. 2016a).

2.6.5 Desenvolvimento de ferramenta

O desenvolvimento de ferramenta de coleta de dados (*crawler*) para redes sociais é facilitado pelo fato de as redes sociais serem bem estruturadas. Os objetos ou indivíduos que figuram a rede social possuem identificações únicas, permitindo dessa forma um maior controle sobre os dados coletados (Benevenuto 2010).

2.6.6 Aplicativos para Facebook

Os Aplicativos Sociais são ferramentas adequadas para serem utilizadas no estudo do comportamento dos usuários e na coleta dos dados do Facebook (Benevenuto 2010).

A arquitetura básica de execução de um Aplicativo Social para Facebook pode ser conforme a Figura 11. A rede social possui um servidor responsável pela comunicação entre o cliente (usuário) e o servidor onde a aplicação processa os seus dados (Benevenuto 2010, Nazir et al. 2008).

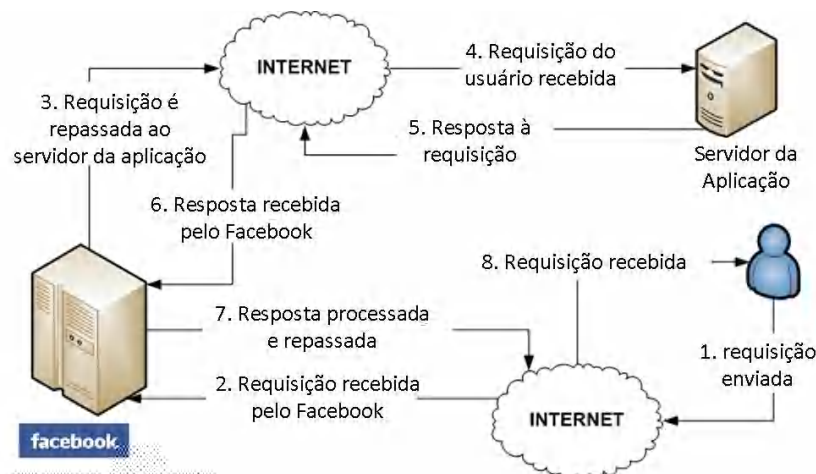


Figura 11 – Arquitetura dos Aplicativos Sociais do Facebook

2.7 Pré-processamento dos dados coletados

Os dados produzidos pelos AS de monitoramento de atividades físicas e disponibilizados nos perfis do Facebook podem conter pequenas inconsistências ou informações incompletas. Isso pode ocorrer: i.) pelo fato de cada AS ter um comportamento próprio em relação a quais dados são disponibilizados e de que forma eles são disponibilizados; ii.) falha de comunicação do AS no momento de gerar a postagem no Facebook, criando uma postagem com informações incompletas; iii.) Erros provocados por falhas dos bio-sensores durante o monitoramento da atividade física, relatando valores incorretos ou incompletos e; iv.) Erros de utilização do AS pelo usuário, registrando tipos de atividades diferentes da executada, por desatenção do usuário no momento de registro da atividade.

Para resolver essas inconsistências é necessária a utilização de técnicas de limpeza de dados, para o preenchimento de informações faltantes, bem como para a

correção de dados discrepantes e a remoção de dados com ruídos. Uma das técnicas simples e eficientes de limpeza de dados é o preenchimento automático de algum atributo inconsistente, através da técnica de *binning* (Han et al. 2011), de acordo com o valor médio do mesmo atributo. Esse preenchimento pode ser feito conforme as características inerentes à situação daquela atividade física como, por exemplo, o usuário que praticou e as atividades físicas com propriedades similares (mesmo tipo de atividade do usuário, mesma distância percorrida ou mesma duração da atividade).

2.8 Classificação dos dados da Rede Social Facebook

Depois do pré-processamento dos dados coletados, pode ocorrer a classificação dos dados, onde o tratamento desses dados classificados possibilitam diferentes análises de seu conteúdo, como o estudo dos PCU, ou identificar interesses e preferências dos usuários. Para facilitar essa tarefa, é importante a identificação dos atributos relevantes na classificação dos dados, os quais descrevem itens nomeados, como pessoas, produtos, atividades executadas, localidades e as informações temporais desses itens (Benevenuto et al. 2012, Paschoal Jr et al. 2016a).

2.8.1 Classificação de dados de padrões comportamentais

O Facebook possui uma extensa lista de AS que trabalham com saúde e *fitness* (Facebook 2014c). Esses aplicativos sociais, quando utilizados pelo usuário, são associados ao perfil do mesmo. Dessa forma, além das publicações feitas por esses aplicativos no mural de notícias do usuário, é possível também consultar a relação de aplicativos que estão associados ao perfil do usuário (Facebook 2014d).

2.9 Processamento Paralelo e Distribuído

Atualmente, a mineração e classificação de dados das RSO representam um problema de *BigData*, pois o volume de dados analisados heterogêneos é enorme e consome cada vez mais recursos computacionais (Jagadish et al. 2014). Assim, essa seção ilustra as atuais tecnologias com alto poder de processamento de grande volume de dados (*BigData*), em um tempo de processamento bem reduzido, comparado com o

processamento em um computador convencional. Essas tecnologias disponibilizam um ambiente altamente escalável, sendo adequadas para a execução do *Workflow* de classificação de dados desta tese.

2.9.1 *Cluster* Computacional

Um *Cluster* computacional é formado pelo agrupamento de nós computacionais, onde esses nós estão interconectados por uma rede de comunicação de dados com alto desempenho. A latência deve ser baixa e os nós podem trabalhar de forma distribuída e com processamento em paralelo para acelerar o tempo de execução de uma atividade computacional (Ogasawara 2011).

Os nós interconectados podem trabalhar em paralelo e representar um único computador com vários núcleos (Ogasawara 2011). O tamanho de um *cluster* computacional pode chegar a centenas ou milhares de computadores (Dean e Ghemawat 2008).

2.9.2 *MapReduce*

O *MapReduce* surgiu com a necessidade da Google Inc. processar, em um tempo reduzido, os seus dados para a manutenção dos serviços disponibilizados aos seus usuários e clientes. Como esses dados são muito volumosos, o seu processamento deve ser feito de forma distribuída em agrupamentos formados com diversos computadores, conhecidos como *clusters* computacionais (Dean e Ghemawat 2008).

Porém, essa tarefa era dificultada, antes da utilização de *MapReduce*, pela complexidade envolvida em uma arquitetura de processamento paralelo e distribuído, incluindo o gerenciamento da distribuição dos dados de forma uniforme para todos os computadores que formavam o *cluster*, o tratamento das falhas de processamento dos dados ocorridos nos computadores e até mesmo a complexidade de programação para contornar essas dificuldades (Dean e Ghemawat 2008).

Para minimizar essas dificuldades foi criada uma abstração, inclusa em uma biblioteca computacional, em que não seria mais necessária a preocupação com esses problemas da arquitetura de processamento paralelo e distribuído, como a tolerância a falhas, a distribuição dos dados e o balanceamento de carga (Dean e Ghemawat 2008).

Essa abstração foi inspirada nos recursos de *Map* e *Reduce* disponibilizados pelo Lisp e por outras linguagens funcionais (Dean e Ghemawat 2008).

O *MapReduce* trabalha em um *cluster* computacional formado por um único nó servidor mestre (*Master*) e diversos nós servidores trabalhadores (*worker*). O *Master* faz a intermediação entre o usuário e a abstração do *MapReduce*, permitindo a interação do usuário de forma que ele submeta as tarefas de *MapReduce*. O *Master* define quais serão os nós *worker* que executarão a função de *Map* e quais serão os nós *worker* que executarão a função de *Reduce*. Ao final do processamento, o *Master* disponibiliza a saída resultante para o usuário. A Figura 12 ilustra o funcionamento da abstração de *MapReduce* (Dean e Ghemawat 2008).

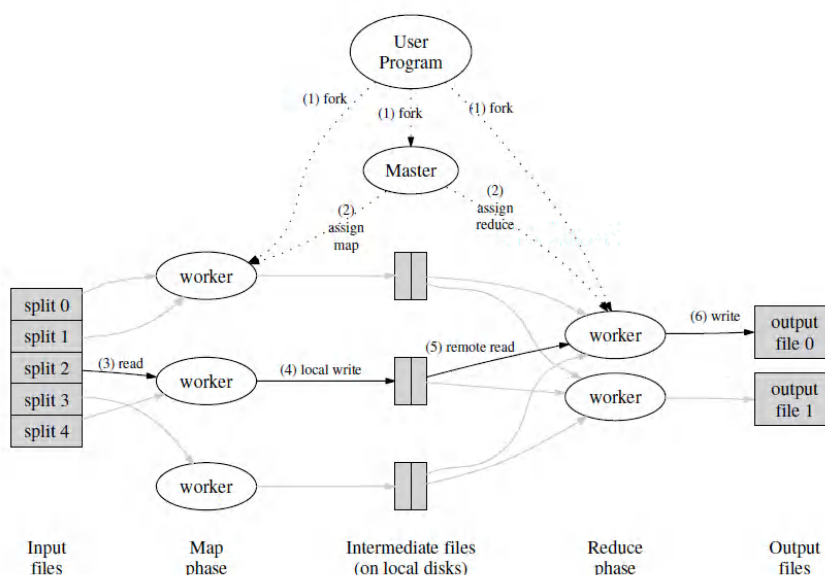


Figura 12 – Funcionamento do *MapReduce*

Esse procedimento é iniciado com a submissão das tarefas do usuário ao *cluster* computacional, identificado como “*User Program*” (Dean e Ghemawat 2008).

O nó *Master* faz com que os dados de entrada dessas tarefas do usuário sejam automaticamente divididos (*splits*), de forma balanceada, entre os nós *worker* que executam a função de *Map*. Essas divisões possuem o tamanho entre 16 *MegaBytes* (MB) e 64 MB, conforme ajuste de parâmetros feito pelo usuário (Dean e Ghemawat 2008).

O nó *Master*, além de ter a sua cópia das tarefas do usuário ((1) *fork*), faz também cópias adicionais ((1) *fork*) dessas tarefas em alguns nós *worker*. Assim, o *Master* faz a escolha dos nós *worker* que executarão a função de *Map* ((2) *assign map*) e

dos nós *worker* que executarão a função de *Reduce* ((2) *assign reduce*) (Dean e Ghemawat 2008).

Cada nó *worker* com a função de *Map* processará os blocos de dados (*split*) associados para o *worker* pelo *Master* ((3) *read*), seguindo o processamento conforme a função de *Map* definida pelo usuário, criando os pares de chaves e valores intermediários em sua memória (Dean e Ghemawat 2008).

A cada período de tempo os nós *worker* com a função de *Map* armazenam os pares de chaves e valores intermediários, que estão em sua memória, em partições de discos locais do *cluster* computacional ((4) *local write*), informando ao nó *Master* a localização dessas partições com os pares de chaves e valores, onde o nó *Master* informa a localização dessas partições a cada nó *worker* com a função de *Reduce* (Dean e Ghemawat 2008).

Os nós *worker* com a função de *Reduce* recebem a notificação do nó *Master* indicando a localização da partição com os dados de *Map* e então executam Chamadas de Procedimento Remoto (RPC – *Remote Procedure Call*) para ler os dados de *Map* dessas partições ((5) *remote read*). É feita a leitura de todos os dados dessas partições e, posteriormente, os pares de chaves e valores intermediários são ordenados de forma que os pares de chaves e valores intermediários idênticos fiquem agrupados (Dean e Ghemawat 2008).

Após essa ordenação e agrupamento, os nós *worker* com a função de *Reduce* processam esses pares de chaves e valores intermediários conforme a função de *Reduce* definida pelo usuário, criando o arquivo de saída final do processamento em uma partição de *reduce* ((6) *write*) (Dean e Ghemawat 2008).

Quando todas as atividades de *Map* e *Reduce* finalizarem, o nó *Master* informa isso ao programa do usuário e o processamento de *Map Reduce* é encerrado (Dean e Ghemawat 2008).

2.9.3 Apache Spark

O Apache Spark é uma plataforma de processamento paralelo e distribuído que estende o modelo de *MapReduce*. Ele estende tanto com mais tipos de processamentos de informações quanto por processar as informações diretamente na memória do *cluster* computacional, ao invés de processar nos discos desse *cluster*. Isso proporciona uma

eficiência muito maior em relação ao *MapReduce* que é executado com o processamento das informações em discos (Karau et al. 2015)

Em comparação com Apache Hadoop (The Apache Software Foundation 2017), o Apache Spark pode executar programas de *MapReduce* com o processamento das informações em memória, com uma velocidade de execução de até cem vezes mais rápida em relação a mesma execução de programas com o processamento das informações em disco feito pelo Hadoop (Apache Spark 2017a). Além disso, pode executar programas de *MapReduce* com o processamento das informações em disco com uma velocidade de execução de até dez vezes mais rápida em relação a mesma execução de programas com o processamento das informações em disco feito pelo Hadoop (Apache Spark 2017a).

O Apache Spark possibilita a criação de *scripts* de execução paralela com o uso de APIs que disponibilizam comandos das linguagens Java, Scala, Python, *Structured Query Language* (SQL) e R. Tem integração para executar programas com o Hadoop e também de acessar os dados do Hadoop (Apache Spark 2017a, Karau et al. 2015).

A Figura 13 ilustra a hierarquia de APIs do Apache Spark. O Apache Spark Core é o núcleo principal responsável pelo gerenciamento do processamento em paralelo no *cluster* computacional. Ele possui as APIs integradas de forma eficiente, trazendo os benefícios de utilizar as características das diferentes APIs, de forma combinada, em um mesmo *script* de execução. As APIs integradas são o Spark SQL (permite análises com dados estruturados), o Spark Streaming (permite análises com o uso de *streaming*), o MLib (permite análises com o uso de *Machine Learning*) e o GraphX (permite análises com o uso de grafos) (Apache Spark 2017a, Karau et al. 2015).

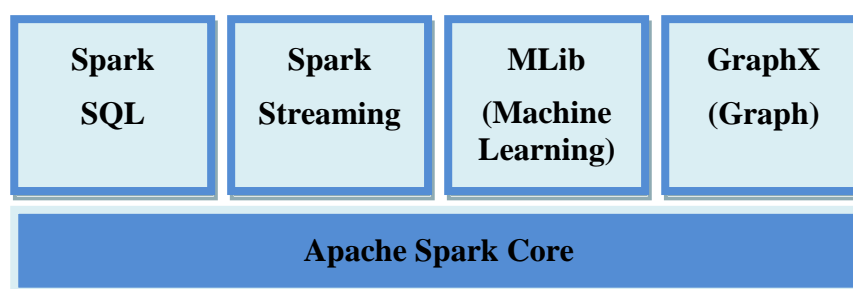


Figura 13 – Hierarquia do Apache Spark

2.9.4 *Resilient Distributed Datasets (RDDs)*

Uma RDD é formada por uma coleção de itens que podem ser distribuídos em um *cluster* computacional e que possibilitam o processamento paralelo e distribuído com tolerância a falhas. O Apache Spark Core possui diversos recursos para criar e manipular as RDDs, incluindo a utilização nas RDDs das funcionalidades de suas APIs (Karau et al. 2015).

As RDDs podem ser criadas a partir de coleções de dados paralelizáveis, como sequências, listas, etc., bem como podem ser criadas a partir de conjunto de dados externos. Elas são passíveis de sofrerem operações de ações e de transformações. Uma ação retorna um valor em resposta a algum processamento com a RDD, como uma operação de *reduce*. Uma transformação cria uma nova RDD como o resultado de um processamento de uma RDD, como uma operação de *map*. O Spark oferece algumas variações de *MapReduce* (Apache Spark 2017b).

As transformações da RDD são executadas de forma tardia, onde o Spark faz o registro das transformações e as executam quando uma ação solicitar o resultado das transformações. Dessa forma, o Spark faz otimizações de execuções das transformações das RDDs (Apache Spark 2017b).

As principais transformações com RDDs compreendem as operações de *map*, *filter*, *flatMap*, *mapPartitions*, *union*, *intersection* e *join*, dentre outras operações. As principais ações com RDDs compreendem as ações de *reduce*, *collect*, *count*, *countByKey* e *foreach*, dentre outras ações (Apache Spark 2017b).

2.9.5 Apache Spark *DataSet/DataFrame*

O *DataSet* é uma interface disponibilizada pelo Apache Spark SQL que evoluiu a partir *SchemaRDD*. Ela é composta por uma coleção de dados distribuída e permite a execução das operações de ações e transformações nesses dados (Apache Spark 2017c). As transformações geram novos *DataSets* e as ações retornam o resultado de algum processamento sobre o *DataSet* (Apache Spark 2017d).

O *DataSet* contém um plano lógico de execução das tarefas para gerar os dados desejados. Esse plano lógico é otimizado e é produzido o plano físico de execução

otimizada para o ambiente paralelo e distribuído. Assim, o *DataSet* é executado de forma otimizada e com um melhor desempenho (Apache Spark 2017d).

O *DataFrame* é um *DataSet* organizado em linhas e colunas, com *schema* definido, tendo um comportamento similar a uma tabela de um banco de dados relacional. Com isso, é possível manipular os seus dados com acesso facilitado aos valores das colunas do *DataFrame*, aplicando as operações e recursos disponibilizados pelo Spark (Apache Spark 2017c, 2017d).

2.9.6 Apache Spark SQL

O Apache Spark SQL é uma biblioteca do Spark, em forma de API, que possibilita o acesso de dados estruturados, ou semi-estruturados, de forma eficiente. São disponibilizadas pelo Spark SQL informações complementares dos dados estruturados manipulados, facilitando o processamento dos mesmos. São também disponibilizadas otimizações complementares, para a execução de operações de ações e de transformações dos dados (Apache Spark 2017e).

Os dados são manipulados na forma de um *SchemaRDD* (versões anteriores do Spark) ou *DataSet/DataFrame* (versões mais recentes do Spark), que é uma RDD com recursos adicionais onde cada linha representa um registro da tabela de dados. Há também *schema* dos dados, que identifica os campos e os seus tipos de dados. Com o *schema*, essa RDD trabalha de forma mais eficiente e permite também a execução de consultas SQL ou *Hive Query Language* (HQL), onde esta última é uma variação de SQL para o Apache Hive. Uma consulta SQL ou HQL feita no *SchemaRDD/DataSet/DataFrame* produz um novo *SchemaRDD/DataSet/Dataframe* como o resultado de retorno da consulta. (Karau et al. 2015).

Os dados manipulados pelo Spark SQL são originados a partir do suporte dos principais formatos de dados estruturados, incluindo JSON e *Comma-Separated Values* (CSV), dentre outros formatos (Apache Spark 2017e, Karau et al. 2015).

2.9.7 Álgebra de *Workflows* Científicos

Segundo a abstração de Álgebra de *Workflows* Científicos (Ogasawara 2011), é possível aplicar a mesma a um conjunto de *DataFrames*, em conjunto com suas operações. Um

DataFrame pode ser representado, através dessa álgebra de *Workflow*, como relações. As operações realizadas sobre os *DataFrames* podem ser representadas, através dessa álgebra de *Workflow*, como atividades que possuem uma determinada semântica para a produção e o consumo dos dados, incluindo as características de paralelismo.

Nessa abstração, as atividades são baseadas em operações algébricas que viabilizam as transformações dos dados, manipulando tuplas de dados, tanto na entrada quanto na saída do *Workflow*. Para isso, são disponibilizadas seis operações, as quais são *Map*, *SplitMap*, *Reduce*, *Filter*, *SRQuery* e *MRQuery*. Nessa álgebra, a razão do consumo de dados pelo produto de tuplas ocorre da seguinte forma: *Map* (1:1), *SplitMap*(1:m), *Reduce* (n:1), *Filter* (1:(0-1)), *SRQuery*(n:m) e *MRQuery* (n:m) (Ogasawara 2011).

Como consequência, é possível executar um *pipeline* de operações com essa álgebra de *Workflow*, onde o *DataFrame* resultante de uma operação é consumido em seguida por outra operação subsequente.

Capítulo 3 - Metodologia

Este capítulo apresenta a metodologia que é proposta por esta tese como um *Framework* com *Workflows*. Ele é sintetizado na Figura 14, indicando o seu processo de análise para se chegar ao PCU. A primeira etapa consiste em proceder as Coletas dos Dados dos perfis dos usuários do Facebook. Esses dados coletados produzem Eventos relacionados aos usuários, os quais indicam as propriedades dos usuários relevantes associadas ao conteúdo coletado.

Esses Eventos são submetidos às Sínteses, as quais produzem Panoramas com os agrupamentos dos Eventos associados a algum intervalo de tempo. Os Panoramas são submetidos às Análises, as quais transformam e resumem os mesmos com os Comportamentos, ou seja, os PCU.

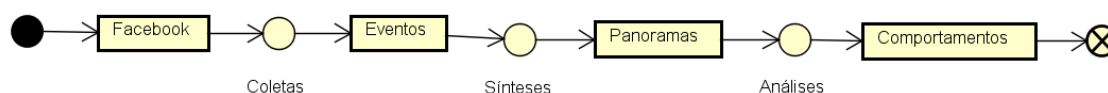


Figura 14 – Processos de Análises

Essa base de proveniência pode ser sintetizada a partir do Diagrama de Classes ilustrado na Figura 15. Ele é subdividido em três blocos ou etapas. Na primeira etapa são consideradas as ações dos usuários no Facebook. Os *Usuários* fazem os *Compartilhamentos* em seus perfis, onde esses *Compartilhamentos* descrevem a ocorrência de eventos genéricos com o registro de alguma determinada observação. Esses *Compartilhamentos* são associados aos *Usuários*. Dentre esses *Compartilhamentos*, alguns são produzidos pelos *Aplicativos Sociais*, relatando propriedades adicionais do usuário e que são passíveis das análises e das classificações desta tese. Dentre os *Aplicativos Sociais*, podem figurar AS de monitoramento de atividades físicas e AS de Coleta de Dados, entre outras possibilidades. Os Aplicativos Sociais são associados ao perfil do Usuário. Os AS de Coleta de Dados são associados aos *Aplicativos Sociais*.

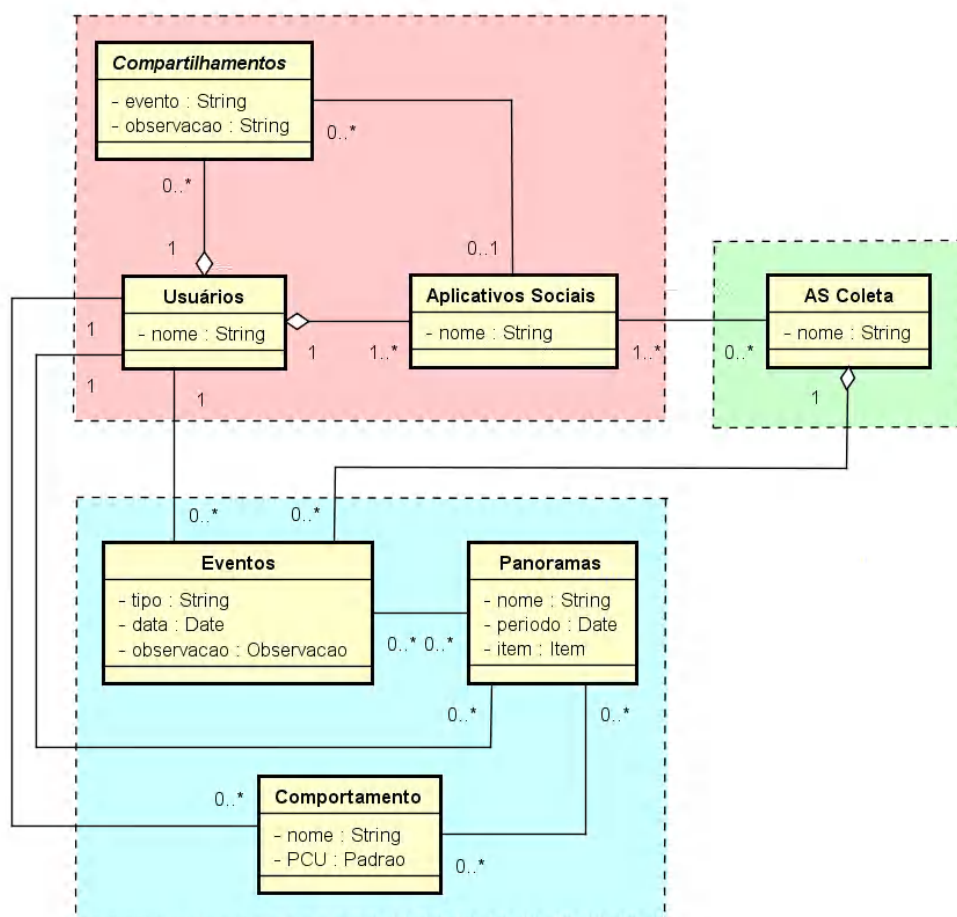


Figura 15 – Diagrama de Classes do Framework

Na segunda etapa são consideradas as ações dos *Usuários* no arcabouço computacional de Coleta dos Dados, relatado neste *Framework* como *AS Coleta*, o qual foi desenvolvido nesta tese. Esse arcabouço permite que os *Usuários* obtenham o resumo das informações coletadas dos dados de seus *Compartilhamentos* feitos pelos *Aplicativos Sociais*, com o foco nos AS de monitoramento de atividades físicas. Esses dados coletados são passíveis de análises e classificações, os quais criam uma base de dados relacionada à prática de atividades físicas. O arcabouço também produz *Compartilhamentos* nos perfis dos *Usuários*, descrevendo novos *Eventos* e observações com sínteses focadas à prática de atividades físicas. O arcabouço é associado aos *Aplicativos Sociais*.

Na terceira etapa são consideradas as ações realizadas no arcabouço computacional de Análise do Padrão Comportamental dos *Usuários* desenvolvido nesta tese. Esse arcabouço permite definir o *Comportamento* dos *Usuários* relatando as ocorrências de *Eventos* associados aos *Usuários*, incluindo o tipo de evento, o que foi

observado e a data de ocorrência desses *Eventos*, obtidos a partir da base de dados criada pelo arcabouço de Coleta dos Dados.

Nessa terceira etapa os *Panoramas* são produtos das transformações/agregações dos *Eventos*. Eles também sofrem atividades de transformações e de Análises, já ilustradas previamente na Figura 14, as quais são relacionadas com os mapeamentos e as reduções dos dados. Assim, os *Panoramas* são formados pelas sínteses do uso dos aplicativos sociais em um determinado intervalo de tempo, com a dinâmica de evolução dos *Panoramas* sendo monitorada pelo modelo de análise proposto nesta tese, na forma de um *pipeline* de operações com os *Panoramas*, estabelecendo o *Comportamento* dos *Usuários*. Dessa maneira, as transformações dos *Panoramas* sintetizam os dados de forma adequada para a realização das Análises que executam as classificações dos dados e estabelecem o Padrão Comportamental do Usuário.

De forma resumida, o *Framework* desta tese estabelece os arcabouços computacionais de Coleta dos Dados e de Análise do Padrão Comportamental dos Usuários. O arcabouço de Coleta dos Dados é um AS de Coleta dos Dados que se integra ao Facebook, faz a coleta dos dados focados nos AS de monitoramento de atividades físicas e permite o compartilhamento de *rankings* competitivos para estimular o combate ao sedentarismo. O arcabouço de Análise do Padrão Comportamental dos Usuários faz a análise dos dados coletados para classificar e estabelecer o comportamento do usuário.

A seguir, é apresentada a arquitetura do *Framework* desta tese, na qual são descritas as etapas de execução das atividades, incluindo as diferentes formas de processamento realizado pelos arcabouços produzidos por esta tese.

Posteriormente, é apresentada a aplicação do arcabouço de Coleta dos Dados no cenário de *fitness*. Esse arcabouço foi denominado FitRank (FitRank 2017a, 2017b). O FitRank tem como atrativo a criação de *rankings* unificados personalizáveis, nos perfis dos usuários, das atividades físicas compartilhadas pelos usuários e por seus amigos do Facebook (Paschoal Jr et al. 2015b, 2016a, 2016b). Finalmente, são apresentadas as características da prática de atividade física, incluindo o modelo de classificação.

3.1 Arquitetura do *Framework*

A arquitetura do *Framework* é sintetizada na Figura 16. Os Perfis de Usuários estão

disponibilizados para a Coleta dos Dados nos servidores do Facebook. Com o acesso pela Internet, o Arcabouço de Coleta dos Dados faz a coleta das atividades físicas compartilhadas por seus usuários em seus perfis.

A execução dessa coleta não é feita de forma paralela. Ela ocorre tanto no momento em que usuários acessam o Arcabouço de Coleta dos Dados, coletando as atividades mais recentes, quanto com a execução de atividades pré-programadas frequentes de coleta de dados complementares (percurso percorrido) ou de atividades mais antigas.

O *Cluster* hospeda os arcabouços computacionais desta tese. A base de dados criada pelo arcabouço de Coleta dos Dados é processada de forma paralelizada e distribuída no *Cluster*, para classificar e gerar as análises dos dados. Após essas análises, o resultado final é a disponibilização dos PCU. O arcabouço de Coleta dos Dados é detalhado na próxima seção e o arcabouço de Análise do Padrão Comportamental dos Usuários é detalhado no Capítulo 4 - Avaliação Experimental desta tese. A disponibilização dos PCU é feita através do acesso ao *Cluster*.

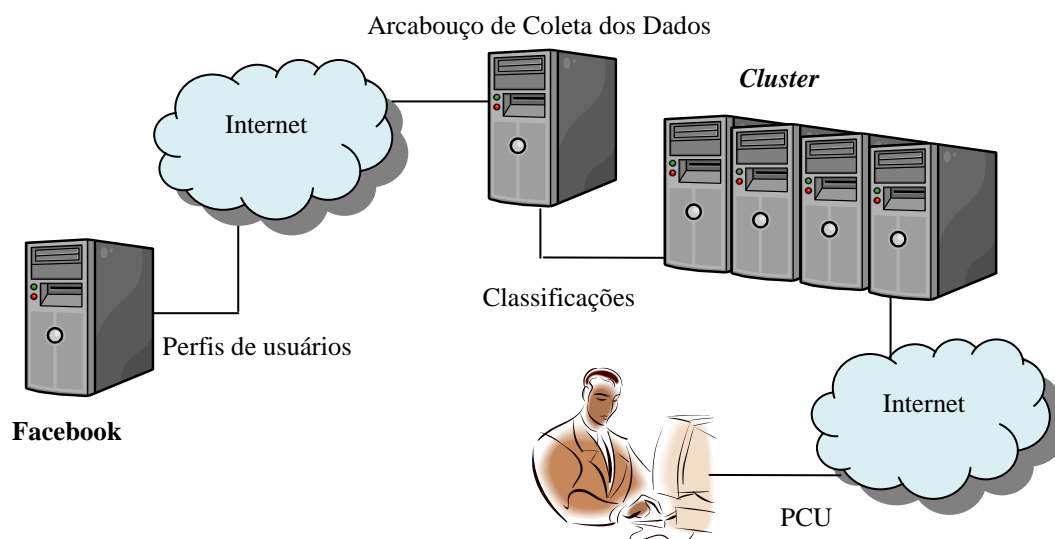


Figura 16 – Arquitetura do Framework

3.2 Aplicação no cenário de *fitness*

No arcabouço de Coleta dos Dados, a coleta de dados é feita com base em padrões de postagens de usuários. Um padrão pode ser expresso por regras, fórmulas ou funções (Freitas et al. 2008a). Para a composição desses padrões, são considerados os AS de

monitoramento de atividades físicas que são aderentes ao *Framework* de classificação desta tese e que permitem a postagem em perfis. Como a relação de AS é flexível, esses padrões devem se adequar à criação de novos AS que sejam aderentes ao *Framework* de classificação desta tese. Para isso, há a necessidade de revisão periódica da relação de aplicativos e dos padrões em uso.

O conjunto de padrões utilizado no arcabouço inclui as postagens feitas pelos AS de monitoramento de atividades físicas disponibilizados na Central de Aplicativos do Facebook (Facebook 2014a), no formato de estória, contendo uma postagem do tipo atividade física. O arcabouço é acessado em <http://eic.cefet-rj.br/app/FitRank/> ou <https://www.facebook.com/fitrank.go/>.

O arcabouço é um AS de coleta de dados do Facebook, o qual tanto coleta os dados de seus usuários quanto também cria novos conteúdos nos perfis dos mesmos no Facebook. Os dados das postagens feitas pelos AS de monitoramento de atividades físicas são coletados pelo arcabouço, tanto do usuário quanto de seus amigos que usam também o arcabouço, possibilitando a criação de *rankings* personalizáveis unificados. Esses *rankings* podem ser compartilhados nos perfis dos usuários do Facebook, possibilitando o aumento da interação entre os usuários. Os compartilhamentos feitos por esse arcabouço, ou seja, os *rankings* compartilhados, não são coletados pelo arcabouço. O estudo dos *rankings* compartilhados é feito diretamente com a análise dos dados do banco de dados deste arcabouço (Paschoal Jr et al. 2016a).

Devido às restrições do Facebook sobre o volume de dados que pode ser coletado de uma vez, a coleta dos dados depende do acesso do usuário ao arcabouço. Isso significa que o processo de coleta ocorre quando o usuário acessa o arcabouço e gera um *ranking*. Nesse momento são coletados os compartilhamentos das atividades físicas do usuário e de seus amigos, relativos aos compartilhamentos que ainda não foram coletados. Caso o usuário ou seus amigos tenham muitas atividades físicas a serem coletadas ou muitas informações complementares de percursos percorridos, essa coleta é feita de forma parcial, a cada acesso ao arcabouço, e também por um agendamento automático de um *job* para a execução de coleta complementar dos dados. Este último ocorre sem a necessidade do usuário acessar novamente o arcabouço.

O principal atrativo do arcabouço é ser um AS que cria uma competição social saudável entre os seus usuários e entre os amigos de seus usuários. Dessa forma, os *rankings* personalizáveis unificados são criados a partir da totalização e resumo das atividades físicas feitas por seus usuários, as quais são corridas, ou caminhadas, ou

ciclismo, ou as três atividades juntas. A personalização do *ranking* pode ser feita com os ajustes de quais atividades serão consideradas para definir a classificação dos usuários, quais critérios de ordenação serão utilizados, como distância percorrida, ou a velocidade média, ou o número de atividades executadas, além de definir o período de tempo que será considerado no *ranking*, como as atividades do dia, da semana, do mês, do ano ou todas as atividades postadas pelos usuários (FitRank 2017a, Paschoal Jr et al. 2016a).

A unificação do *ranking* do arcabouço ocorre com a coleta das atividades físicas postadas por diferentes AS de monitoramento de atividades físicas, onde os usuários que utilizam distintos AS de monitoramento de atividades físicas estarão classificados no mesmo *ranking* unificado criado pelo arcabouço. Tanto a unificação do *ranking* quanto a personalização dos parâmetros de classificação do *ranking* são fatores inovadores e inéditos criados pelo arcabouço, onde essas características não foram disponibilizadas por outros AS, antes da criação do arcabouço (Paschoal Jr et al. 2016a).

Um dos principais objetivos de concepção e de utilização do arcabouço é poder motivar e estimular a queda do sedentarismo humano. Para isso, os usuários podem postar os seus *rankings* unificados personalizáveis, gerados pelo arcabouço, em seus perfis no Facebook. Essa inovação poderá proporcionar uma maior motivação dos usuários e de seus amigos à prática de atividades físicas. Assim, eles poderão passar a ter um comportamento mais saudável (Paschoal Jr et al. 2016b).

A Figura 17 ilustra o esquema de funcionamento do arcabouço de Coleta dos Dados. Inicialmente, o usuário faz a prática de sua atividade física com o monitoramento de seu AS de atividades físicas preferido. Ao término dessa atividade física, o usuário faz a postagem da mesma em seu perfil do Facebook, por meio do compartilhamento da mesma disponível em seu AS de monitoramento de atividades físicas, no formato de postagem de atividade física. Quando esse usuário ou os amigos dele fizerem o próximo uso do arcabouço, o arcabouço fará a coleta dos dados dessa atividade física monitorada compartilhada no Facebook, com o auxílio das APIs RestFB e Graph API, e armazenará esses dados no banco de dados do arcabouço. Posteriormente, quando o usuário solicitar a criação de algum *ranking* unificado personalizável, é feito o processamento dos dados com suas análises para a criação do respectivo *ranking*. Se usuário desejar, ele pode fazer a postagem de seu *ranking* em seu perfil do Facebook (Paschoal Jr et al. 2015b, 2016a).

O arcabouço trabalha atualmente com a coleta dos dados de um conjunto de dezesseis AS de monitoramento de atividades físicas (Paschoal Jr et al. 2015b, 2016a), que são: Endomondo Corrida Ciclismo MTB (Endomondo 2017a, 2017b), MapMyFitness (MapMyFitness 2017a, 2017b), MapMyRide Equitação Ciclismo (MapMyRide 2017a, 2017b), MapMyRun (MapMyRun 2017a, 2017b), MapMyWalk (MapMyWalk 2017a, 2017b), Nike+ Running (Nike 2017a, 2017b), Polar Beat Fitness Coach (Polar Beat 2017a, 2017b), Polar Flow Activity & Sports (Polar Flow 2017a, 2017b), RunKeeper GPS Correr e Caminhar (RunKeeper 2017a, 2017b), Runtastic Corrida e Caminhada (Runtastic 2017a, 2017b), Runtastic PRO GPS Correr (Runtastic 2017c), Runtastic Road Bike Ciclismo (Runtastic 2017d), Runtastic Road Bike PRO GPS (Runtastic 2017e), Runtastic Mountain Bike GPS (Runtastic 2017f), Runtastic Mountain Bike PRO (Runtastic 2017g) e Strava GPS Correr Ciclismo (Strava 2017a, 2017b). A adição da coleta de dados de novos AS de atividades físicas é feita através de revisões periódicas de evolução de versão do arcabouço (Paschoal Jr et al. 2016a).

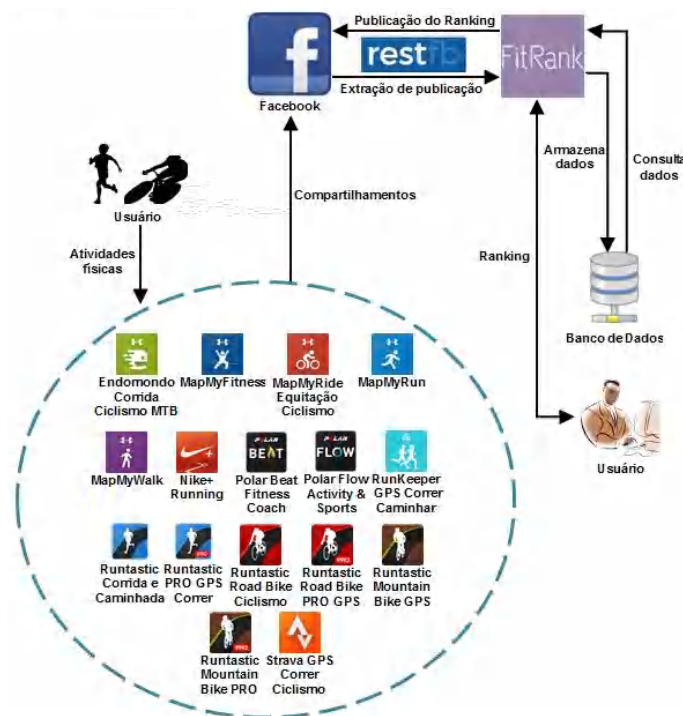


Figura 17 – Esquema de funcionamento do Arcabouço de Coleta dos Dados

A maioria dos AS de monitoramento de atividades físicas faz a postagem da atividade física realizada pelo usuário como uma estória estruturada (Facebook 2017c) ou como uma estória mais rica e personalizada (Facebook 2016a). Isso facilita a coleta de dados, pelo fato de a postagem ter a identificação do AS que a gerou, bem como um resumo da execução da atividade física e, opcionalmente, o geo-referenciamento da

mesma. A Figura 18 ilustra postagens feitas por alguns AS de monitoramento de atividades físicas, as quais são passíveis de coleta dos dados caso esses usuários passem a utilizar o arcabouço.

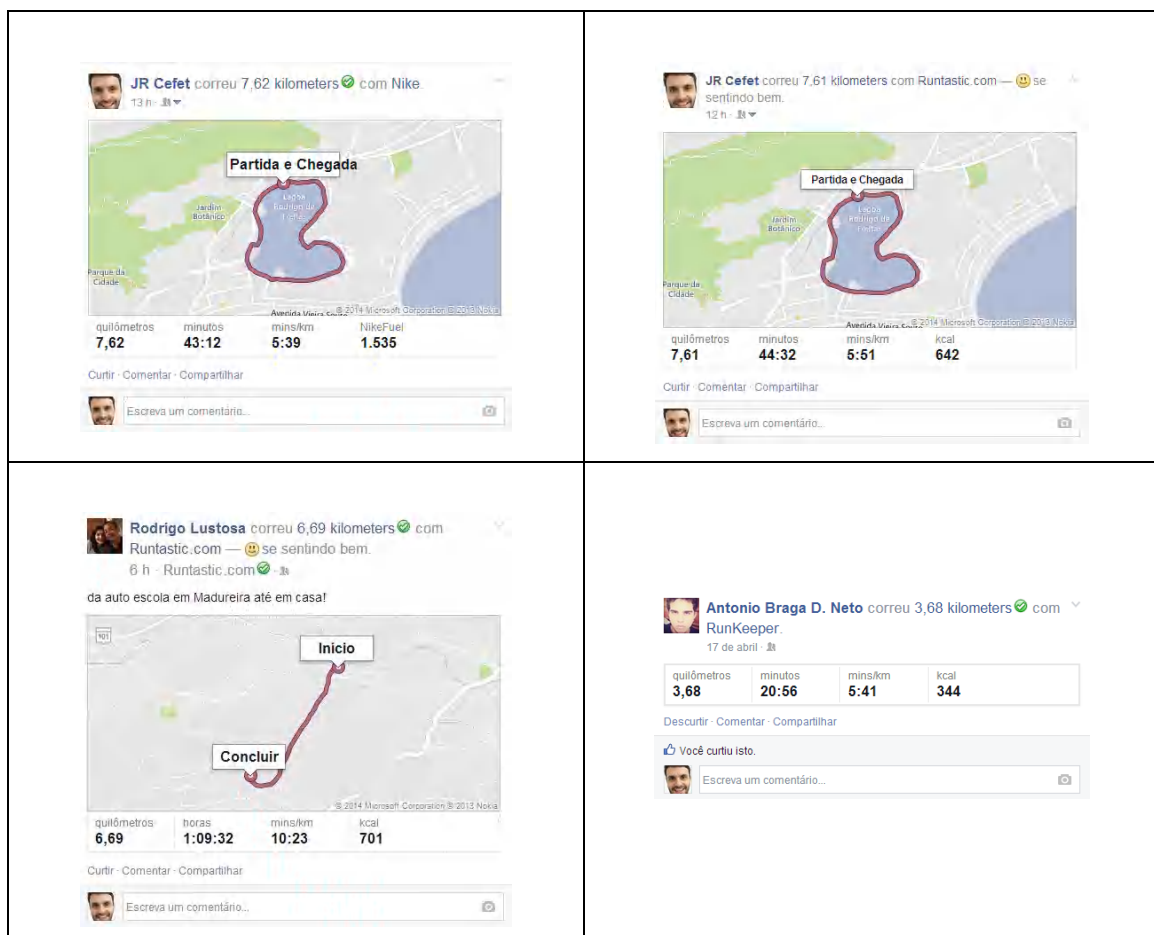


Figura 18 – Postagens de AS de monitoramento de atividades físicas

A Figura 19 ilustra dois *rankings* reais unificados personalizados de um usuário do arcabouço, bem como a postagem de um desses *rankings* no perfil deste usuário do arcabouço no Facebook. O *ranking* unificado personalizado com as atividades físicas postadas no perfil do usuário com as suas corridas, caminhadas e ciclismo realizados no último mês é ilustrado na Figura 19(a). O *ranking* unificado personalizado com todas as atividades físicas postadas no perfil do usuário com as suas corridas, caminhadas e ciclismo é ilustrado na Figura 19(b). O *ranking* ilustrado pela Figura 19(b) é postado no perfil do usuário no Facebook pelo arcabouço conforme ilustrado na Figura 19(c) (Paschoal Jr et al. 2016a).

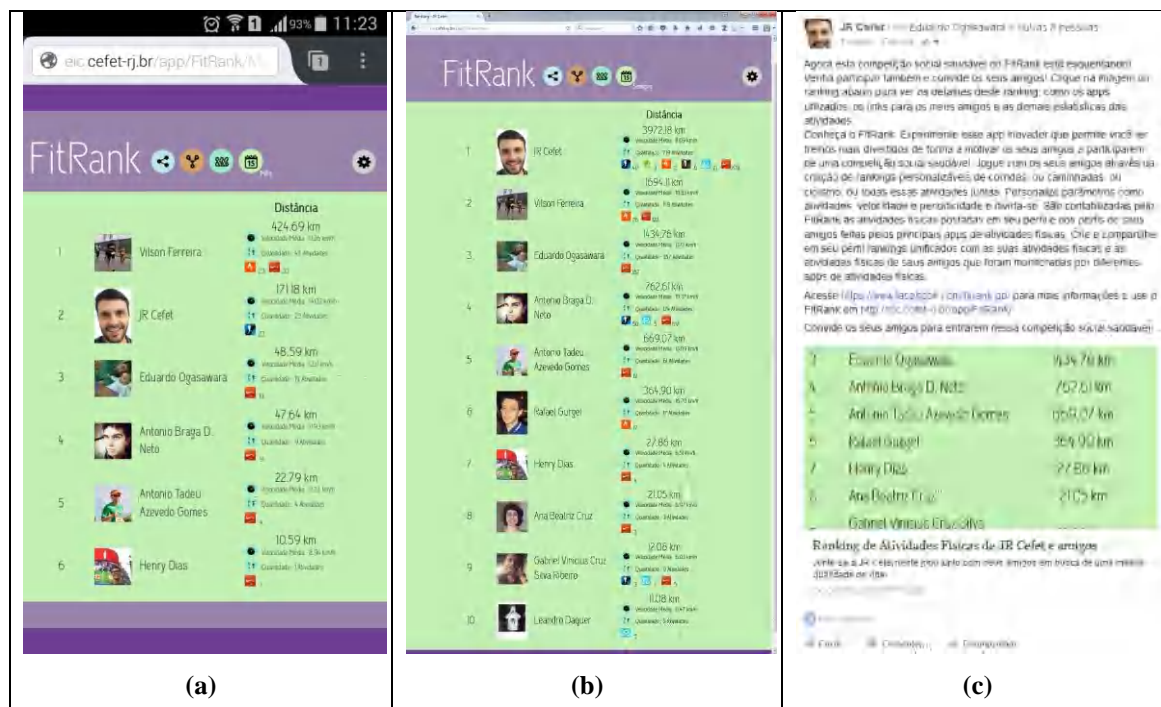


Figura 19 – Rankings de usuário do arcabouço com a sua postagem no Facebook

3.3 Sedentarismo x Atividades Físicas

O sedentarismo é um dos maiores problemas de saúde pública da atualidade. Na vida moderna, muitas facilidades e inovações tecnológicas são criadas pela humanidade, as quais muitas vezes simplificam as atividades manuais, podendo até mesmo reduzi-las drasticamente ou até mesmo as eliminando completamente (Paschoal Jr et al. 2016a).

Essas facilidades e inovações são benéficas e possibilitam que o indivíduo tenha mais tempo livre com a sua família e ao seu bem-estar. Contudo, o cuidado com a sua saúde e com o seu corpo devem estar incluídos em seu bem estar, mas, infelizmente, esse cuidado está aquém do ideal (Paschoal Jr et al. 2016a).

Para tentar mudar um pouco essa triste realidade, um objetivo secundário desta tese é estimular os usuários dos arcabouços computacionais desta tese a combaterem o sedentarismo com a ajuda da tecnologia e do apelo das RSO. Para isso, a definição do PCU é uma forma de avaliar a atual situação do usuário e poderá fornecer ao mesmo uma análise que o estimule à mudança de seu estilo vida. Assim, ele poderá ter uma vida mais saudável, incluindo uma boa qualidade de vida, se preocupando com os cuidados com a saúde e combatendo o sedentarismo com o incentivo da prática regular de atividades físicas, o que é recomendado pela *World Health Organization* – (WHO) (Paschoal Jr et al. 2016a, World Health Organization 2002).

É perturbador o fato de que 25% da população adulta mundial tenha um comportamento sedentário. E é mais preocupante ainda que 80% da população adolescente mundial também tenha um comportamento sedentário. Conforme a WHO, os indivíduos sedentários têm uma maior probabilidade de risco de morte, por volta de 20% a 30%, em comparação com os indivíduos que praticam regularmente as atividades físicas (Paschoal Jr et al. 2016a, World Health Organization 2010, 2017).

3.4 Atividades físicas como prevenção de doenças

Segundo a WHO, a falta da prática regular de atividades físicas, por parte dos indivíduos, é um dos dez maiores fatores de risco de morte. Além disso, também é um fator de risco para a ocorrência de diabetes, problemas cardiovasculares e câncer (Paschoal Jr et al. 2016a, World Health Organization 2017).

Portanto, a prática regular de atividades físicas é fundamental na prevenção desses e de outros problemas de saúde, bem como proporciona uma melhor qualidade de vida aos indivíduos. Os benefícios dessa prática regular, conforme sugere a WHO (Paschoal Jr et al. 2016a, World Health Organization 2010, 2017), são: i.) melhora do controle de peso; ii.) redução de doenças coronárias; iii.) redução de depressão; iv.) redução do câncer de cólon; v.) redução da hipertensão; vi.) melhora da saúde óssea; vii.) melhora do condicionamento cardiorespiratório; viii.) redução de diabetes; ix.) fortalecimento dos músculos; x.) redução do risco de quedas; xi.) melhora do equilíbrio de energia do organismo; xii.) redução do risco de fratura de vértebras; xiii.) redução do câncer de mama e; xiv.) redução de derrames.

3.5 Modelo de classificação de comportamento

Como forma de combate ao sedentarismo, a WHO recomenda a prática de atividades físicas de acordo com o que está relacionado na Tabela 1 (Paschoal Jr et al. 2016a, World Health Organization 2010, 2017). Essa recomendação foi o modelo utilizado na classificação do comportamento saudável do usuário proposto nesta tese, ou seja, ela foi o modelo que definiu o PCU, onde foram inferidos aos usuários os atributos **“(Pouco Saudável – PS)”**, **“(Saudável – S)”** e **“(Muito Saudável – MS)”**.

Tabela 1 – Recomendações da WHO sobre a prática de atividades físicas

Faixa Etária	Recomendações da WHO
De 5 a 17 anos	<ul style="list-style-type: none"> • Ao menos 60 minutos de atividades físicas moderadas ou vigorosas por dia. (Saudável) • Atividades físicas com mais de 60 minutos por dia fornecem benefícios adicionais à saúde. (Muito Saudável) • Ao menos 3 vezes por semana atividades físicas que fortaleçam músculos e ossos. (Muito Saudável)
De 18 a 64 anos	<ul style="list-style-type: none"> • Ao menos 150 minutos de atividades físicas moderadas ou 75 minutos de atividades físicas vigorosas por semana. (Saudável) • 300 minutos de atividades físicas moderadas por semana fornecem benefícios adicionais à saúde. (Muito Saudável) • Ao menos 2 vezes por semana atividades físicas que fortaleçam músculos ou que envolvam uma quantidade significativa de músculos. (Muito Saudável)
A partir de 65 anos	<ul style="list-style-type: none"> • Ao menos 150 minutos de atividades físicas moderadas ou 75 minutos de atividades físicas vigorosas por semana. (Saudável) • 300 minutos de atividades físicas moderadas por semana fornecem benefícios adicionais à saúde. (Muito Saudável) • Ao menos 3 vezes por semana atividades físicas para melhorar o equilíbrio e prevenir quedas. (Muito Saudável) • Ao menos 2 vezes por semana atividades físicas que fortaleçam músculos ou que envolvam uma quantidade significativa de músculos. (Muito Saudável)

Portanto, a classificação do PCU foi feita conforme a sua faixa etária. Assim, o comportamento **“Pouco Saudável” (PS)** foi inferido quando o usuário não atingiu o valor mínimo indicado como **“Saudável” (S)**, incluindo as situações de excessos da prática de atividades físicas. O comportamento **S** foi inferido quando o usuário atingiu o mínimo indicado como **“(Saudável)”**. O comportamento **“Muito Saudável” (MS)** foi inferido quando o usuário atingiu o mínimo indicado como **“(Saudável)”** e também ao menos um dos itens indicados com **“(Muito Saudável)”** (Paschoal Jr et al. 2016a).

Capítulo 4 - Avaliação Experimental

A avaliação experimental desta tese buscou fazer a avaliação do *Framework* proposto, no sentido de avaliar os PCU no contexto da prática de atividades físicas. Nesse sentido, esse capítulo está organizado de forma que avaliou: i.) a modelagem dos *Workflows*; ii.) a identificação dos PCU e; iii.) a execução do *Workflow* de modo paralelo e distribuído. Ao final, é apresentada uma discussão dessa avaliação experimental.

Para a realização da avaliação experimental, foi feita a instanciação dos arcabouços da metodologia proposta nesta tese, na forma de *Workflows*, e já ilustrada previamente na Figura 15. Assim, o “AS Coleta” foi instanciado como o arcabouço de Coleta dos Dados. Os “Eventos”, “Panoramas” e “Comportamento” foram instanciados pelo arcabouço de Análise do Padrão Comportamental dos Usuários, incluindo as atividades de Transformações e de Análises, para a definição dos PCU.

A seção “3.2 Aplicação no cenário de *fitness*” desta tese detalhou o funcionamento da instanciação representada pelo arcabouço de Coleta dos Dados. Os dados coletados sofreram a limpeza dos dados com a aplicação da técnica de *binning* (Han et al. 2011), que corrigiu as informações ausentes ou incompatíveis com a modalidade da atividade física compartilhada e coletada. A correção foi feita com base na média de execução de atividades com características similares. Foram exemplos de aplicação dessa técnica corrida com 24 horas de duração, corrida com velocidade média de 40 km/h, ciclismo com velocidade média de 70 km/h, caminhada com velocidade de 10 km/hora, dentre outras situações. É importante destacar que o uso dessa técnica não pode ser generalizado em dados de saúde oriundos da coleta de bio-sensores, devendo ser analisado os riscos inerentes ao tipo de dado a ser corrigido. O volume de dados corrigidos foi relativamente baixo, por volta de 1,8% das atividades físicas coletadas.

4.1 Modelagem dos *Workflows*

A avaliação experimental desta tese foi feita com *Workflows* para a identificação dos PCU com o uso dos AS de monitoramento de atividades físicas para a classificação do comportamento saudável do usuário. Para isso, foram analisadas as postagens criadas pelos AS de monitoramento de atividades físicas, compartilhadas no perfil do usuário. Elas permitem definir os hábitos de execução das atividades físicas, incluindo as suas

propriedades, disponíveis nas informações do perfil do usuário, na seção de atividades físicas³ (Paschoal Jr et al. 2016a).

Essas análises foram feitas com base nas informações armazenadas dos usuários e de suas atividades físicas executadas, inclusive as informações de geo-referenciamento, de bio-sensoriamento, do AS de monitoramento de atividades físicas utilizado e também informações temporais da atividade física (Paschoal Jr et al. 2016a).

Para a definição do PCU de comportamento saudável, foram consideradas como principais informações a serem classificadas a data de execução da atividade, a distância percorrida, a velocidade, a duração, a quantidade de atividades físicas, a idade e o sexo. A intensidade da atividade física (leve, moderada ou vigorosa) foi inferida conforme a modalidade (corrida, caminhada e ciclismo) correlacionada com a velocidade (Paschoal Jr et al. 2016a). Essa correlação foi baseada no modelo do Compêndio de Atividades Físicas, que relaciona as atividades físicas como caminhar, correr e pedalar com o nível de esforço de execução da atividade física, ou seja, a intensidade da atividade física (Coelho-Ravagnani et al. 2013, Farinatti 2003).

4.1.1 Ambiente computacional da avaliação experimental

Para o processamento dos dados desta tese, foi utilizado o *cluster* computacional da Escola de Informática & Computação (EIC) do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ), o qual disponibilizou o Apache Spark com as suas APIs. Este *cluster* computacional é ilustrado na Figura 20.

O arcabouço de Coleta dos Dados está hospedado no servidor de aplicações “albali”. O arcabouço de Análise do Padrão Comportamental dos Usuários está hospedado no servidor “aldebaran”, com acesso por *Secure Shell* (SSH) e que é o nó principal de servidor do Apache Spark. O processamento paralelo é gerenciado pelo “aldebaran”, com a distribuição de processamento com os demais nós servidores *workers* do Apache Spark. Assim, os servidores nós *workers* do Apache Spark são “kaus”, “rukbat”, “arietis” e “librae”. Cada nó servidor do Apache Spark possui 4 núcleos de processamento, totalizando 20 núcleos. A configuração básica de cada nó consiste em 4 processadores x86_64 Intel Xeon E3-1220 3.1 *Giga Hertz* (GHz), com

³ <https://www.Facebook.com/JRCefet/fitness>

memória *caches* L1, L2 e L3, além de 14.6 *Giga Bytes* (GB) de memória *Random Access Memory* (RAM).

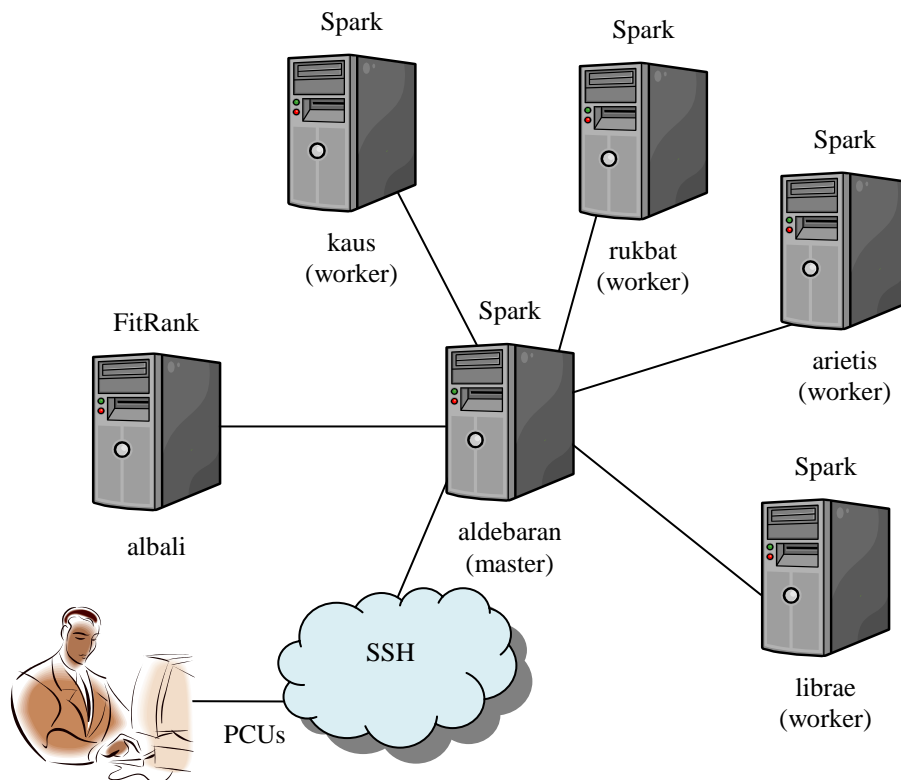


Figura 20 – Cluster da EIC

No arcabouço computacional de Análise do Padrão Comportamental dos Usuários foi utilizado o processamento paralelo e distribuído do Apache Spark. Os dados foram processados com o auxílio do Apache Spark SQL, com o uso de *DataFrames*, com algumas transformações de *MapReduce* de dados e análise desses dados transformados para chegar aos PCUs. Dessa forma, a arquitetura permitiu o escalonamento.

A escolha do Apache Spark se deu pelo motivo de permitir o processamento paralelo e distribuído, com o recurso de *Map* e *Reduce*, de forma otimizada com a execução em memória, além dos recursos disponibilizados por suas APIs integradas.

4.1.2 População e Grupos

A população de usuários deste estudo constituiu, ao todo, 378 usuários do arcabouço de Coleta dos Dados. Essa população foi formada por 292 usuários do sexo masculino (77%) e 86 usuários do sexo feminino (23%). Ela é identificada como “Total” na Tabela

2. As análises desse grupo possibilitaram os resultados obtidos nos estudos dos arcabouços produzidos por esta tese.

Tabela 2 – População e Grupos de Usuários

População e Grupos							
Identificação	Usuários	Grupo	% do Total	% Grupo	Masculinos	% Masculinos	Femininos % Femininos
Total	378	Total	100%	N/A	292	77%	86 23%
Rankings Válidos	125	G1	33%	N/A	98	78%	27 22%
Com atividades antes e depois do FitRank	111	G1-1	29%	89%	83	75%	28 25%
Rankings Inválidos	253	G2	67%	N/A	194	77%	59 23%
Rankings Inválidos com Atividades Físicas	171	G2-1	45%	68%	128	75%	43 25%
Rankings Inválidos sem Atividades Físicas	82	G2-2	22%	32%	63	77%	19 23%

Dessa população de 378 usuários, 253 usuários geraram *rankings* inválidos (G1) e 125 geraram ao menos um *ranking* válido (G2) durante a utilização do arcabouço de Coleta dos Dados.

O critério que definiu se o *ranking* foi válido foi de o usuário ter ajustado os parâmetros corretos de criação de seu *ranking*. Isso significa que ele selecionou a modalidade de atividades do ranking e o período a ser considerado no *ranking* conforme a realidade de execução de suas atividades físicas. Em síntese, um usuário Corredor selecionou um *ranking* de Corrida com o período de tempo que ele tenha praticado alguma corrida.

Um usuário gerou um *ranking* inválido quando ele selecionou uma modalidade que não tinha praticado, como um Ciclista ter gerado um *ranking* de Corrida. E também foi um *ranking* inválido quando o usuário selecionou um período de tempo em que não praticou a atividade física daquela modalidade como, por exemplo, um Ciclista que gerou um *ranking* de pedaladas da última semana e ele tinha feito a sua última pedalada há 15 dias. Assim, um *ranking* inválido teve o valor da atividade física do usuário zerada, pela ausência da prática de atividades físicas com as características configuradas pelo usuário.

Dessa forma, a população total de usuários, de 378, pode ser representada em grupos e subgrupos de usuários de acordo com as características da prática de atividades físicas e com a geração de *rankings* válidos. Esses grupos e subgrupos foram criados e proporcionaram análises com focos específicos e que serão relatados neste capítulo.

O primeiro grupo foi formado pelos usuários que têm atividades físicas compartilhadas e que geraram *Rankings* Válidos (G1), sendo representado por 125 usuários (33% do total de usuários). Esse grupo foi composto por 98 usuários masculinos (78%) e 27 usuários femininos (22%). Ele foi identificado na Tabela 2 como “Ranking Válidos” e como grupo “G1”.

Houve uma subdivisão do “G1”, com 89% da população do “G1”, ou 29% da população total, o qual foi representado com 111 usuários. Ele foi formado por 83 usuários masculinos (75%) e 28 usuários femininos (25%), os quais tiveram a prática de atividades físicas tanto antes quanto após o início do uso do arcabouço de Coleta dos Dados. A sua identificação na Tabela 2 está como “Com atividades antes e depois do FitRank” e como grupo “G1-1”.

O segundo grupo foi formado pelos usuários que geraram apenas *Rankings* Inválidos (G2), sendo representado por 253 usuários (67% do total de usuários). Esse grupo foi composto por 194 usuários masculinos (77%) e 59 usuários femininos (23%). Ele foi identificado na Tabela 2 como “Rankings Inválidos” e como grupo “G2”.

Porém, houve duas subdivisões do “G2”. O primeiro subgrupo foi formado com 68% da população do “G2”, ou 45% da população total, o qual foi representado com 171 usuários. Ele foi formado por 128 usuários masculinos (75%) e 43 usuários femininos (25%), os quais tiveram *rankings* inválidos, mas tinham atividades físicas compartilhadas. A sua identificação na Tabela 2 está como “Ranking Inválidos com Atividades Físicas” e como grupo “G2-1”.

O segundo subgrupo foi formado com 32% da população do “G2”, ou 22% da população total, o qual foi representado com 82 usuários. Ele foi formado por 63 usuários masculinos (77%) e 19 usuários femininos (23%), os quais foram considerados como fora do escopo da tese e não sofreram estudos ou análises. A sua identificação na Tabela 2 está como “Ranking Inválidos sem Atividades Físicas” e como grupo “G2-2”.

Um fator que pode ser sugestionado para a existência da população de usuários do grupo “G2-1” pode ser alguma falha de comunicação da interface do arcabouço de Coleta dos Dados ou da divulgação do mesmo. Essa população teve o perfil de usuário adequado, mas esses usuários podem ter encontrado alguma dificuldade de utilização do arcabouço.

Um fator que pode ser sugestionado para a existência da população de usuários do grupo “G2-2” pode ser também alguma falha de comunicação da interface do arcabouço de Coleta dos Dados ou da divulgação do mesmo, no sentido de que seja esclarecida a forma de utilização do mesmo e o perfil adequado de usuário (que utiliza AS de monitoramento de atividades físicas e que compartilha essas atividades em seu perfil).

A Tabela 3 resumiu e agrupou alguns dos grupos da Tabela 2 o que proporcionou uma análise mais detalhada e com resultados adequados, conforme o arcabouço analisado e as características da análise realizada.

Tabela 3 – População de Usuários e Grupos Estudados

População e Grupos Estudados								
Análises	Usuários	Grupo	% do Total	% Grupo	Masculinos	% Masculinos	Femininos	% Femininos
Uso do arcabouço de Coleta dos Dados	378	Total	100%	N/A	292	77%	86	23%
Uso do arcabouço de Coleta dos Dados	125	G1	33%	N/A	98	78%	27	22%
Uso do arcabouço de Coleta dos Dados	111	G1-1	29%	89%	83	75%	28	25%
Arcabouço de Análise do Padrão Comportamental dos Usuários	296	G1 e G2-1	78%	N/A	226	76%	70	24%

O grupo “Total” foi analisado no uso do arcabouço de Coleta dos Dados, em análises de *rankings* gerados e de suas características. O grupo “G1” foi analisado no uso do arcabouço de Coleta dos Dados, em análises complementares do comportamento do usuário. O grupo “G1-1” foi analisado no uso do arcabouço de Coleta dos Dados, enriquecendo as análises do comportamento do usuário, em específico com o comportamento anterior e posterior ao uso deste arcabouço.

Os grupos “G1” e “G2-1” foram agrupados e analisados no arcabouço de Análise do Padrão Comportamental dos Usuários. Esse agrupamento foi formado com 78% da população total, o qual foi representado com 296 usuários. Ele foi formado por 226 usuários masculinos (76%) e 70 usuários femininos (24%). Assim, essa foi a população de usuários com atividades físicas compartilhadas, que sofreram a coleta desses dados e que tiveram os seus PCU estabelecidos.

4.1.3 Análise do *Workflow* de Coleta dos Dados

A análise desse *Workflow* foi feita com base no uso do arcabouço de Coleta dos Dados, o que permitiu a observação de alguns comportamentos dos usuários. A população considerada inicialmente nessa análise foi o total de usuários do AS de Coleta dos Dados, ou seja, 378 usuários.

A Figura 21(a) ilustra a quantidade de usuários da população, conforme o gênero. Houve a prevalência do gênero masculino. A Figura 21(b) ilustra a quantidade de *rankings* gerados por essa população, conforme a modalidade da atividade física. Foi gerado um total de n=4.691 *rankings*. A modalidade denominada “Misto” representou os *rankings* gerados de forma unificada, considerando as totalizações das execuções das atividades de corrida + caminhada + ciclismo, em conjunto em um mesmo *ranking*. Houve a prevalência da geração de *rankings* de Corridas.

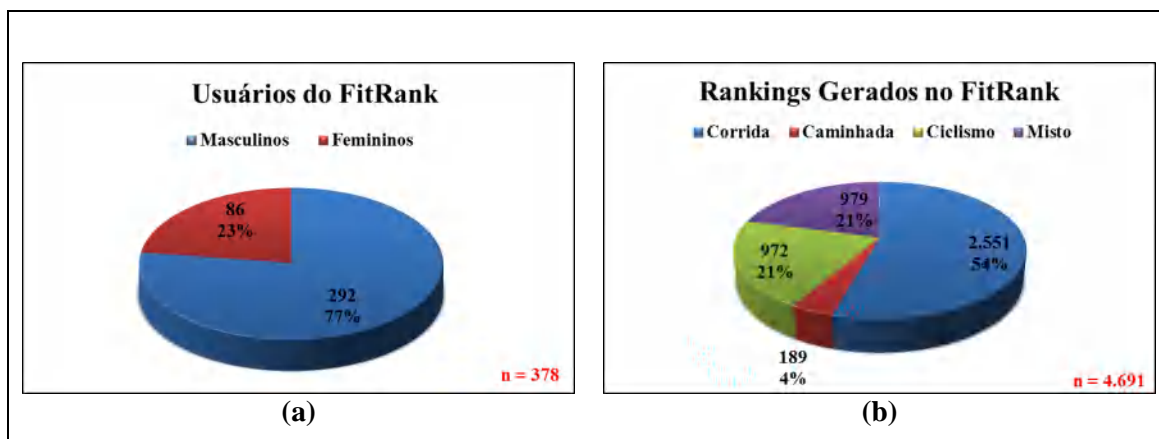


Figura 21 – Modalidades de *Rankings* gerados

A Figura 22(a) ilustra o total de *rankings* gerados em cada modalidade, conforme o gênero. Houve a prevalência da geração de *rankings* de Corridas em ambos os gêneros. A Figura 22(b), porém, ilustra que dessa população de 378 usuários, 253 usuários geraram *rankings* inválidos e 125 geraram ao menos um *ranking* válido.

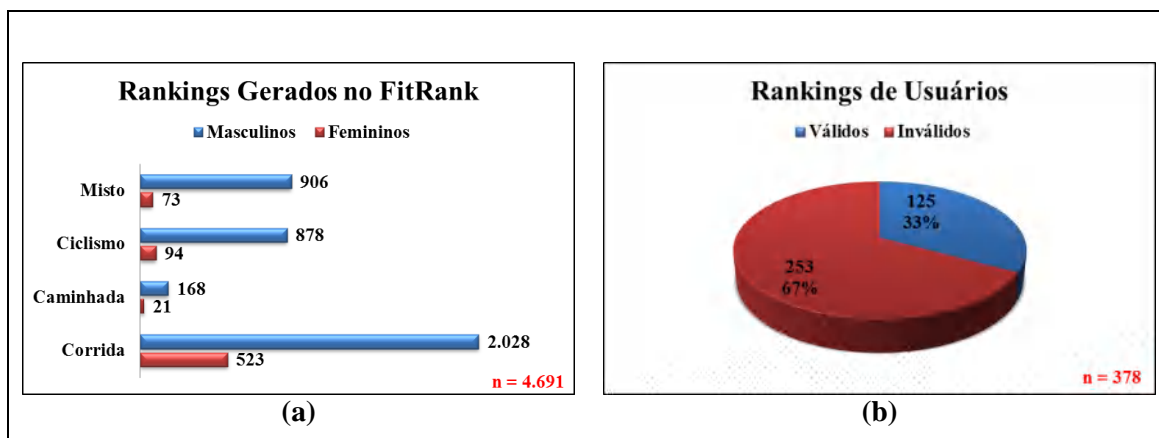


Figura 22 – Modalidades de *Rankings* e Grupos de Usuários

A seguir são apresentados os resultados obtidos com a população de usuários com *rankings* válidos, totalizando n=125 usuários, já que esses resultados foram motivados pelo uso correto do arcabouço de Coleta dos Dados. Esse foi o grupo “G1” descrito anteriormente na seção “4.1.2 População e Grupos” desta tese.

A Figura 23(a) ilustra a quantidade de usuários da população com *rankings* válidos, conforme o gênero. Houve a prevalência do gênero masculino. A Figura 23(b) ilustra a faixa etária dessa população, conforme o gênero. A faixa etária identificada com “0” representou os usuários que especificaram em seus perfis apenas o dia e o mês da data de nascimento, sem terem especificados o ano de nascimento. Os valores indicados como “não especificado” representaram os usuários que não disponibilizaram

em seus perfis essa informação. Houve a prevalência dos usuários masculinos na faixa etária entre de 18 a 39 anos. Os usuários femininos tiveram a prevalência na faixa etária de 40 a 60 anos. Alguns usuários não especificaram essa informação em seus perfis ou não preencheram o ano de nascimento, em ambos os gêneros.

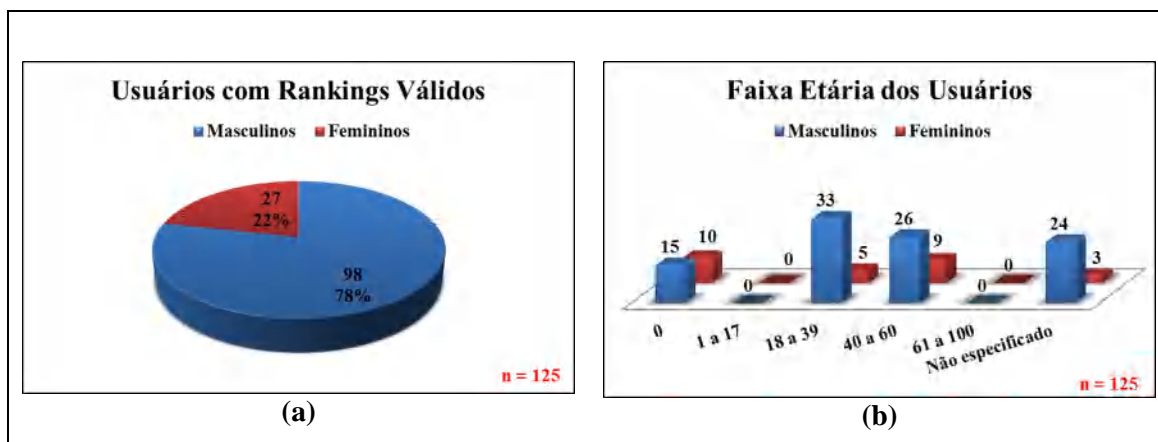


Figura 23 – Usuários com *Rankings* Válidos e Faixa Etária

A Figura 24(a) ilustra as regiões geográficas desses usuários, conforme o gênero. Houve a prevalência de usuários na região Sudeste do Brasil, em ambos os gêneros. Houve a ocorrência de usuários de outros países, no gênero masculino. Alguns usuários não especificaram essa informação em seus perfis. A Figura 24(b) ilustra a escolaridade desses usuários, conforme o gênero. Houve a prevalência do nível de escolaridade superior, em ambos os gêneros. Alguns usuários não especificaram essa informação em seus perfis.

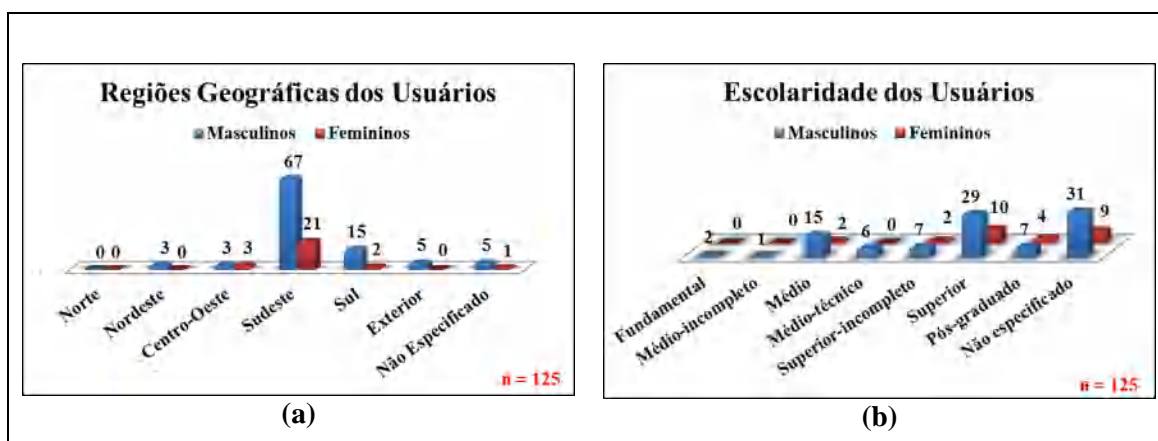


Figura 24 – Regiões Geográficas e Escolaridade dos Usuários

A Figura 25(a) ilustra o estado civil desses usuários, conforme o gênero. Houve a prevalência de usuários casados no gênero masculino e de usuários solteiros no gênero feminino. Alguns usuários não especificaram essa informação em seus perfis. A Figura

25(b) ilustra a quantidade máxima de pessoas nos *rankings* desses usuários, conforme o gênero. Houve a prevalência de 2 a 5 pessoas nos *rankings* dos usuários em ambos os gêneros. Isso sugestionou que esses usuários tiveram até 4 amigos figurados no mesmo *ranking*, sugestionando também um bom nível de sociabilidade dentre os mesmos, no uso do arcabouço de Coleta dos Dados.

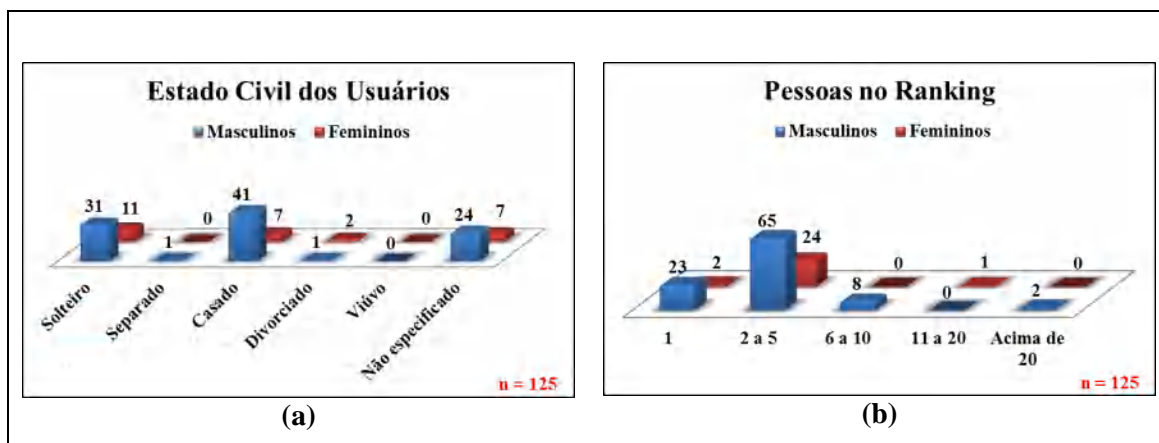


Figura 25 – Estado Civil e Pessoas no *Ranking* dos Usuários

A Figura 26(a) ilustra as diferentes configurações utilizadas nas criações dos *rankings* desses usuários, conforme o gênero. Houve a prevalência de usuários que fizeram de 2 a 5 ajustes de diferentes configurações na geração de seus *rankings*, em ambos os gêneros, sugestionando que eles exploraram algumas das diferentes variações de *rankings*. A Figura 26(b) ilustra a quantidade de *rankings* gerados por esses usuários, conforme o gênero. Houve a prevalência de 1 a 10 *rankings* gerados, em ambos os gêneros. Isso sugestionou que esses usuários eram recentes na utilização do arcabouço de Coleta dos Dados.

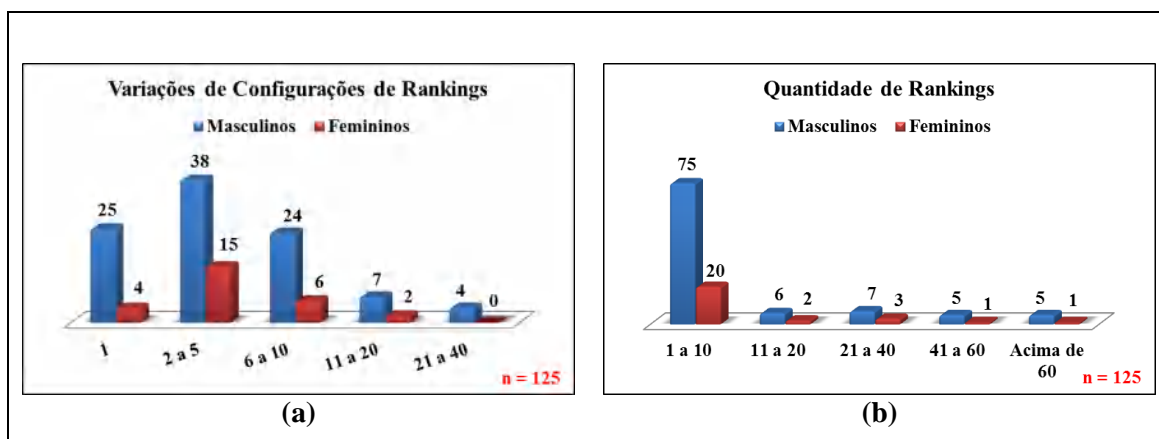


Figura 26 – Configurações e Quantidade de *Ranking* dos Usuários

A Figura 27(a) ilustra as diferentes modalidades de *rankings* gerados por esses usuários, conforme o gênero. Eles geraram um total de $n=2.572$ *rankings*. Houve a

prevalência da geração de *rankings* de corridas, em ambos os gêneros.

A Figura 27(b) ilustra o estudo feito com a população de usuários do grupo “G1-1”, descrito anteriormente na seção “4.1.2 População e Grupos” desta tese, representada pela quantidade de n=111 usuários. Esses usuários criaram *rankings* válidos e tinham atividades físicas compartilhadas tanto antes quanto depois da data de início de utilização do arcabouço de Coleta dos Dados. Assim, nessa figura temos o histórico de compartilhamento de atividades físicas desses usuários em seus perfis, conforme o gênero. Houve a prevalência de usuários que compartilharam as suas atividades físicas há pelo menos 25 meses, com o limite de até 48 meses, em ambos os gêneros. Isso sugestionou que esses usuários compartilhavam suas atividades físicas em seus perfis há um bom tempo.

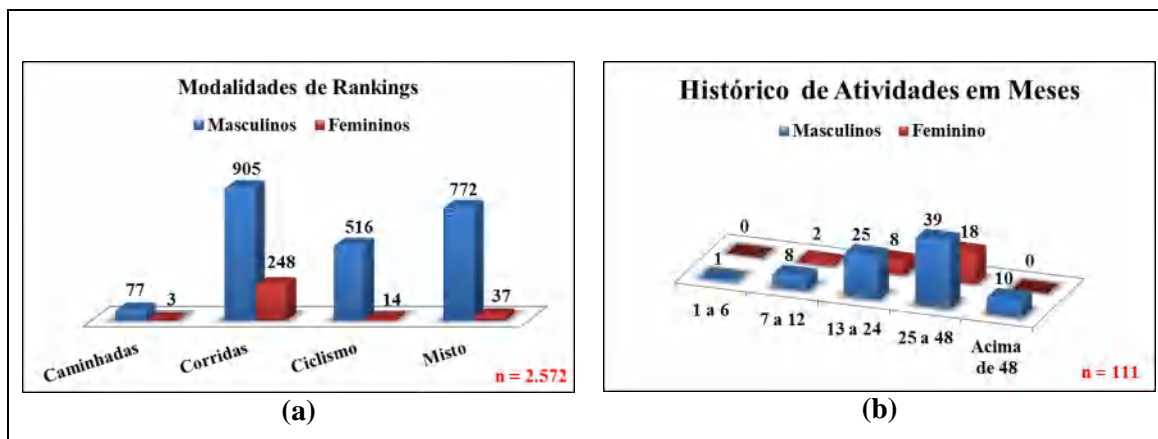


Figura 27 – Modalidades de Ranking e Histórico de Atividades dos Usuários

A Figura 28(a) ilustra o tempo de utilização do arcabouço de Coleta dos Dados por esses usuários, conforme o gênero. Houve a prevalência do tempo de utilização de até 6 meses, em ambos os gêneros. Isso sugestionou que a maioria desses usuários iniciou a utilização deste arcabouço havia pouco tempo. A Figura 28(b) ilustra a quantidade de atividades físicas compartilhadas por esses usuários em seus perfis, conforme o gênero. Houve a prevalência de usuários masculinos que compartilharam de 101 a 200 atividades físicas e de usuários femininos que compartilharam de 1 a 50 atividades físicas. Houve uma distribuição significativa de usuários nas outras faixas de atividades físicas, com o limite de até 400 atividades físicas compartilhadas. Isso sugestionou que esses usuários compartilharam uma quantidade significativa de atividades físicas.

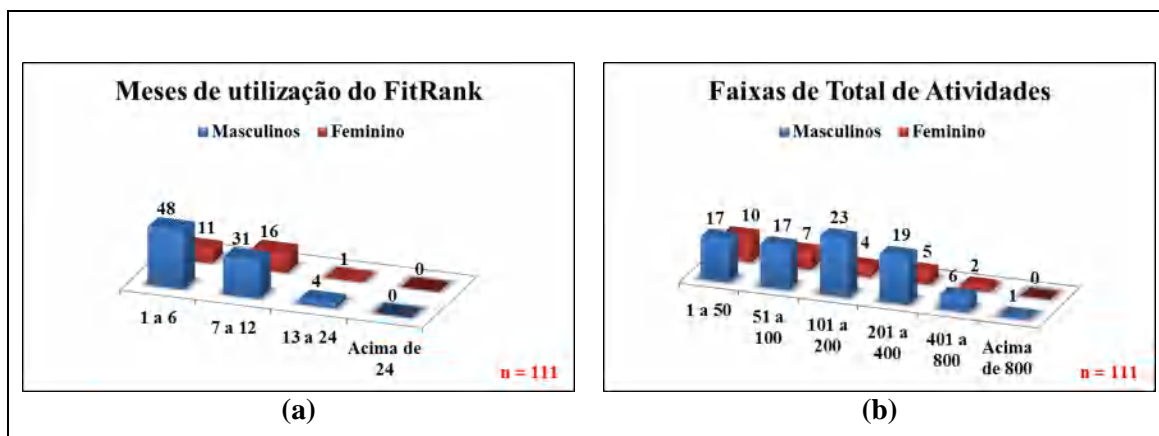


Figura 28 – Tempo de utilização do FitRank e Total de Atividades dos Usuários

Em resumo, foi sugestionado o comportamento padrão do usuário no arcabouço de Coleta dos Dados, com base nos itens mais frequentes desta população. Dessa forma, ele pode ter sido do gênero masculino, ter gerado uma quantidade maior de *rankings* na modalidade de Corrida, ter residido na região Sudeste do Brasil e com o seu nível de escolaridade superior. Ele pode ter tido até 4 amigos em seu *ranking*, sugestionando um bom nível de interação com seus amigos que usaram este arcabouço. Ele pode ter feito até 5 ajustes diferentes de criação de *rankings*, pode ter gerado de 1 a 10 *rankings*, pode ter iniciado o compartilhamento de suas atividades físicas no intervalo entre 25 e 48 meses, bem como pode ter iniciado o uso deste arcabouço nos últimos 6 meses.

Em relação aos gêneros, foram sugestionadas como diferenças desta população a faixa etária, o estado civil e a quantidade de atividades físicas compartilhadas, com base nos itens mais frequentes. Assim, o usuário masculino pode ter tido a sua idade entre 18 e 39 anos e o feminino entre 40 a 60 anos. O usuário masculino pode ter sido casado e o feminino solteiro. O usuário masculino pode ter compartilhado de 101 a 200 atividades físicas e o feminino de 1 a 50 atividades físicas.

4.1.4 *Workflow* de Análise Comportamental

Agora será detalhado o funcionamento do *Workflow* representado pelo arcabouço de Análise do Padrão Comportamental dos Usuários. A base de dados do arcabouço de Coleta de Dados tem seu acesso disponibilizado no servidor “alderaban” do *cluster* da EIC, já ilustrado previamente na Figura 20, para o arcabouço de Análise do Padrão Comportamental dos Usuários. Essa base de dados foi processada no Apache Spark

SQL, de forma paralelizada nos 3 nós do Apache Spark totalizando 12 núcleos. Os dados foram processados com o uso de *DataFrames*.

Assim ocorreram as instanciações de “*Eventos*”, “*Panoramas*” e “*Comportamento*”. Os registros das ocorrências das atividades físicas realizadas pelos usuários do arcabouço de Coleta dos Dados foram definidos como *Eventos*. Eles incluem os principais atributos de execução da atividade física, os quais sofreram Transformações de dados e se tornaram *Panoramas* com informações mais detalhadas da prática de atividade física do usuário.

Porém, os *Panoramas* estavam ainda em um nível intermediário de informações que não possibilitavam o estabelecimento dos PCU. Foram necessárias mais Transformações e Análises que reduziram os dados aos *Comportamentos*, e que possibilitaram, finalmente, o estabelecimento do PCU.

O *Workflow* foi operado com a metodologia de álgebra de *Workflow* (Ogasawara 2011), com a utilização de duas operações dessa álgebra. Foram utilizadas as operações de *Map* e *Reduce*, de forma que foi feito um *pipeline* de realização dessas operações, onde o resultado de uma operação foi imediatamente consumido pela próxima operação, de forma paralela e distribuída. Essa álgebra de *Workflow* (Ogasawara 2011) foi escolhida por permitir o *pipeline* de *Map* e *Reduce*.

Cada operação de *Map* fez alguma transformação dos dados que recebeu como entrada. A primeira operação de *Map* teve, como entrada de dados, os dados obtidos a partir da base de dados do arcabouço de Coleta dos Dados, contendo a relação de todas as atividades físicas praticadas pelos usuários. As demais operações de *Map* tiveram, como entrada de dados, os dados resultantes da Transformação anterior feita no *pipeline* de operações do *Workflow*.

Da mesma forma, as operações de *Reduce* também tiveram, como entrada de dados, os dados resultantes da Transformação anterior feita no *pipeline* de operações do *Workflow*.

O primeiro *DataFrame*, denominado **DF_post_fitness**, contém os dados das atividades físicas coletadas e disponibilizadas pelo arcabouço de Coleta dos Dados. Ele continha n=30.391 atividades físicas coletadas.

A Figura 29 ilustra o *schema* de **DF_post_fitness** e a sua quantidade de tuplas. Nele estavam disponibilizadas todas as informações das atividades físicas executadas, bem como as informações básicas iniciais que permitiram as análises dos PCU.

```
scala> DF_post_fitness.printSchema()
root
 |-- id_pessoa: string
 |-- nome: string
 |-- data_publicacao: string
 |-- modalidade: string
 |-- id_app: string
 |-- distancia_percorrida: double
 |-- duracao: double
 |-- aplicativo: string
 |-- temporal: string
 |-- intensidade: string
 |-- MET: double
 |-- velocidade: double
 |-- genero: string
 |-- data_cadastro: string
 |-- dias: long
 |-- semanas: integer

scala> DF_post_fitness.count()
res35: Long = 30391
```

Figura 29 – *Schema* e quantidade de tuplas de *DF_post_fitness*

Assim, a Figura 30 detalha a instanciación das atividades da metodologia, com a identificação dos *Eventos*, *Panoramas*, *Comportamento* e *Análise dos Resultados*. A primeira operação da álgebra de *Workflow* foi um *Map* para criar novos atributos. Ela foi identificada como a atividade **Calculo_Indicadores_Intensidades**, a qual foi regida pelo operador *Map*, que produziu os atributos *semana*, *horasModeradas*, *horasVigorosas*, *minutosModerados* e *minutosVigorosos*. Esses atributos foram criados para identificar, em cada atividade, qual a semana que ela foi praticada pelo usuário, além de identificar se a atividade é moderada ou vigorosa e, nesse caso, a transformação da duração da atividade física em horas e minutos. Assim, a atividade **Calculo_Indicadores_Intensidades** foi representada por *Map*(nome, semana, horasModeradas, horasVigorosas, minutosModerados, minutosVigorosos).

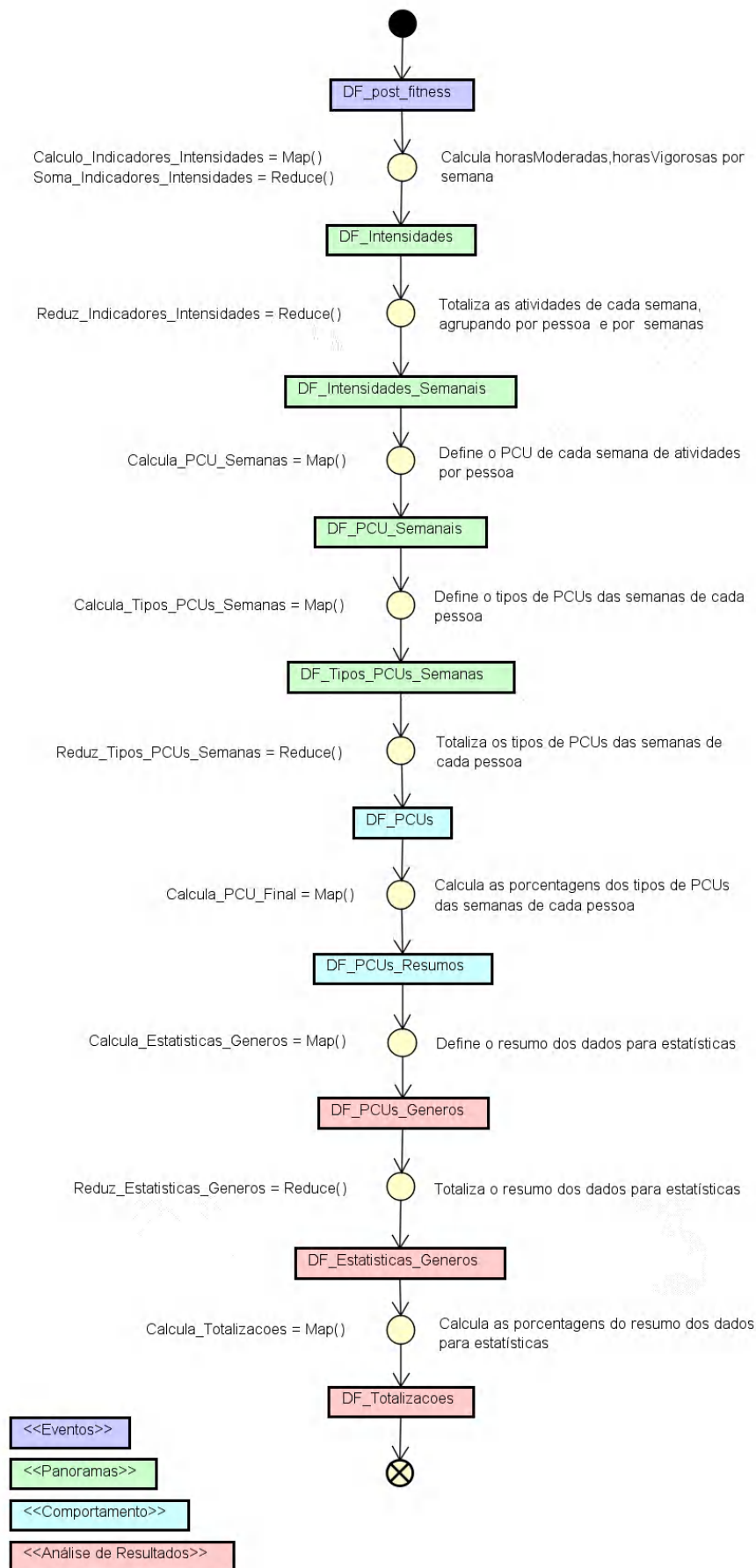


Figura 30 – Workflow da metodologia

O resultado da atividade **Calculo_Indicadores_Intensidades** foi consumido imediatamente como entrada para a segunda operação da álgebra de *Workflow*, identificada como a atividade **Soma_Indicadores_Intensidades**. Essa atividade foi regida pelo operador *Reduce*, que somou os atributos horasModeradas, horasVigorosas, minutosModerados e minutosVigorosos. Essa soma ocorreu totalizando na mesma semana as horasModeradas, as horasVigorosas, os minutosModerados e os minutosVigorosos, mas ainda com a duplicidade de tuplas em uma mesma semana. Assim, tinha a tupla da semana com a redução das atividades moderadas e também a tupla da mesma semana com a redução das atividades vigorosas.

Ao término de execução da atividade **Soma_Indicadores_Intensidades**, o resultado foi o *DataFrame* denominado **DF_Intensidades**, o qual continha n=14.486 semanas de atividades físicas. Assim, a atividade **Soma_Indicadores_Intensidades** foi representada por **Reduce**(nome, semana, horasModeradas, horasVigorosas, minutosModerados e minutosVigorosos). A Figura 31 ilustra o *schema* de **DF_Intensidades** e a sua quantidade de tuplas.

```
scala> DF_Intensidades.printSchema()
root
 |-- id_pessoa: string
 |-- nome: string
 |-- genero: string
 |-- semana: integer
 |-- atividadesSemana: long
 |-- horasModeradas: double
 |-- horasVigorosas: double
 |-- minutosModerados: double
 |-- minutosVigorosos: double

scala> DF_Intensidades.count()
res36: Long = 14486
```

Figura 31 – *Schema* e quantidade de tuplas de **DF_Intensidades**

O resultado da atividade **Soma_Indicadores_Intensidades** foi consumido imediatamente como entrada para a terceira operação da álgebra de *Workflow*, identificada como a atividade **Reduz_Indicadores_Intensidades**. Essa atividade foi regida pelo operador *Reduce*, que uniu em uma mesma tupla as tuplas da mesma semana. Assim, as totalizações das horasModeradas, das horasVigorosas, dos minutosModerados e dos minutosVigorosos de uma mesma semana ficaram em uma única tupla daquela semana do usuário.

Ao término de execução da atividade **Reduz_Indicadores_Intensidades**, o resultado foi o *DataFrame* denominado **DF_Intensidades_Semanais**, o qual continha n=10.657 semanas de atividades físicas. Assim, a atividade **Reduz_Indicadores_Intensidades** foi representada por **Reduce**(nome, semana, horasModeradas, horasVigorosas, minutosModerados e minutosVigorosos). A

Figura 32 ilustra o schema de **DF_Intensidades_Semanais** e a sua quantidade de tuplas.

```
scala> DF_Intensidades_Semanais.printSchema()
root
|-- id_pessoa: string
|-- nome: string
|-- genero: string
|-- semana: integer
|-- atividadesSemana: long
|-- horasModeradas: double
|-- horasVigorosas: double
|-- minutosModerados: double
|-- minutosVigorosos: double

scala> DF_Intensidades_Semanais.count()
res38: Long = 10657
```

Figura 32 – *Schema* e quantidade de tuplas de **DF_Intensidades_Semanais**

O resultado da atividade **Reduz_Indicadores_Intensidades** foi consumido imediatamente como entrada para a quarta operação da álgebra de *Workflow*, identificada como a atividade **Calcula_PCU_Semanas**. Essa atividade foi regida pelo operador *Map*, que criou um novo atributo que definiu o PCU da semana. Então, nesse ponto, cada semana foi classificada com o PCU equivalente. Ao término de execução da atividade **Calcula_PCU_Semanas**, o resultado foi o *DataFrame* denominado **DF_PCUs_Semanais**, o qual continha n=10.657 semanas de atividades físicas. Assim, a atividade **Calcula_PCU_Semanas** foi representada por **Map**(nome, semana, PCUSemana). A Figura 33 ilustra o *schema* de **DF_PCUs_Semanais** e a sua quantidade de tuplas.

```
scala> DF_PCUs_Semanais.printSchema()
root
 |-- id_pessoa: string
 |-- nome: string
 |-- genero: string
 |-- semana: integer
 |-- atividadesSemana: long
 |-- horasModeradas: double
 |-- horasVigorosas: double
 |-- minutosModerados: double
 |-- minutosVigorosos: double
 |-- PCUSemana: string

scala> DF_PCUs_Semanais.count()
res40: Long = 10657
```

Figura 33 – *Schema* e quantidade de tuplas de **DF_PCUs_Semanais**

O resultado da atividade **Calcula_PCU_Semanas** foi consumido imediatamente como entrada para a quinta operação da álgebra de *Workflow*, identificada como a atividade **Calcula_Tipos_PCUs_Semanas**. Essa atividade foi regida pelo operador *Map*, que criou novos atributos, os quais foram semanasAtivas, PS, S e MS. Esses atributos foram criados para identificar, dentre as semanas com os seus PCU, quais foram Poucos Saudáveis (PS), quais foram Saudáveis (S) e quais foram Muito Saudáveis (MS). Ao término de execução da atividade **Calcula_Tipos_PCUs_Semanas**, o resultado foi o *DataFrame* denominado **DF_Tipos_PCUs_Semanas**, o qual continha n=10.657 semanas de atividades físicas com a identificação dos respectivos tipos de PCU. Assim, a atividade **Calcula_Tipos_PCUs_Semanas** foi representada por **Map**(nome, semanasAtivas, PS, S, MS). A Figura 34 ilustra o *schema* de **DF_Tipos_PCUs_Semanas** e a sua quantidade de tuplas.

O resultado da atividade **Calcula_Tipos_PCUs_Semanas** foi consumido imediatamente como entrada para a sexta operação da álgebra de *Workflow*, identificada como a atividade **Reduz_Tipos_PCUs_Semanas**. Essa atividade foi regida pelo operador *Reduce*, que somou os atributos semanasAtivas, PS, S e MS. Essa soma ocorreu totalizando em uma única tupla do usuário o total de semanas ativas de execução de atividades físicas e os totais de semanas com PCU Poucos Saudáveis, Saudáveis e Muito Saudáveis.

```
scala> DF_Tipos_PCUs_Semanas.printSchema()
root
|-- id_pessoa: string
|-- nome: string
|-- genero: string
|-- semanasAtivas: long
|-- PS: long
|-- S: long
|-- MS: long

scala> DF_Tipos_PCUs_Semanas count()
res42: Long = 10657
```

Figura 34 – Schema e quantidade de tuplas de DF_Tipos_PCUs_Semanas

Ao término de execução da atividade **Reduz_Tipos_PCUs_Semanas**, o resultado foi o *DataFrame* denominado **DF_PCUs**, o qual continha n=296 usuários com PCU de atividades físicas. Assim, a atividade **Reduz_Tipos_PCUs_Semanas** foi representada por **Reduce**(nome, semanasAtivas, PS, S, MS). A Figura 35 ilustra o *schema* de **DF_PCUs** e a sua quantidade de tuplas.

```
scala> DF_PCUs.printSchema()
root
|-- id_pessoa: string
|-- nome: string
|-- genero: string
|-- semanasAtivas: long
|-- PS: long
|-- S: long
|-- MS: long

scala> DF_PCUs.count()
res44: Long = 296
```

Figura 35 – Schema e quantidade de tuplas de DF_PCUs

O resultado da atividade **Reduz_Tipos_PCUs_Semanas** foi consumido imediatamente como entrada para a sétima operação da álgebra de *Workflow*, identificada como a atividade **Calcula_PCU_Final**. Essa atividade foi regida pelo operador *Map*, que criou novos atributos, os quais foram porcentagemPS, porcentagemS, porcentagemMS e PCUFinal. Esses atributos foram criados para identificar as respectivas porcentagens de cada tipo de PCU de atividades físicas que o usuário teve. E, a partir disso, estabeleceu o PCUFinal do usuário com base no PCU em que teve uma maior frequência de ocorrência.

Ao término de execução da atividade **Calcula_PCU_Final**, o resultado foi o *DataFrame* denominado **DF_PCUs_Resumos**, o qual continha n=296 usuários com os seus PCU de atividades físicas definidos e estabelecidos. Assim, a atividade **Calcula_PCU_Final** foi representada por *Map*(nome, porcentagemPS, porcentagemS, porcentagemMS, PCUFinal). A Figura 36 ilustra o *schema* de **DF_PCUs_Resumos** e a sua quantidade de tuplas.

```
scala> DF_PCUs_Resumos.printSchema()
root
|-- id_pessoa: string
|-- nome: string
|-- genero: string
|-- semanasAtivas: long
|-- PS: long
|-- porcentagemPS: double
|-- S: long
|-- porcentagemS: double
|-- MS: long
|-- porcentagemMS: double
|-- PCUFinal: string

scala> DF_PCUs_Resumos.count()
res46: Long = 296
```

Figura 36 – *Schema* e quantidade de tuplas de **DF_PCUs_Resumos**

Para fins de resumos estatísticos, o resultado da atividade **Calcula_PCU_Final** foi consumido imediatamente como entrada para a oitava operação da álgebra de *Workflow*, identificada como a atividade **Calcula_Estatisticas_Generos**. Essa atividade foi regida pelo operador *Map*, que criou novos atributos, os quais foram masculinos, femininos, PSMasculinos, SMasculinos, MSMasculinos, PSFemininos, SFemininos e MSFemininos. Esses atributos foram criados para identificar, entre os usuários, quantos masculinos e femininos têm os respectivos PCU Pouco Saudáveis, Saudáveis e Muito Saudáveis.

Ao término de execução da atividade **Calcula_Estatisticas_Generos**, o resultado foi o *DataFrame* denominado **DF_PCUs_Generos**, o qual continha n=296 usuários masculinos ou femininos com PCU estabelecidos. Assim, a atividade **Calcula_Estatisticas_Generos** foi representada por *Map*(masculinos, femininos, PSMasculinos, SMasculinos, MSMasculinos, PSFemininos, SFemininos, MSFemininos). A Figura 37 ilustra o *schema* de **DF_PCUs_Generos** e a sua quantidade de tuplas.

```
scala> DF_PCUs_Generos.printSchema()
root
|-- masculinos: long
|-- femininos: long
|-- PS: long
|-- S: long
|-- MS: long
|-- PSMasculinos: long
|-- SMasculinos: long
|-- MSMasculinos: long
|-- PSFemininos: long
|-- SFemininos: long
|-- MSFemininos: long

scala> DF_PCUs_Generos.count()
res48: Long = 296
```

Figura 37 – *Schema* e quantidade de tuplas de **DF_PCUs_Generos**

O resultado da atividade **Calcula_Estatisticas_Generos** foi consumido imediatamente como entrada para a nona operação da álgebra de *Workflow*, identificada como a atividade **Reduz_Estatisticas_Generos**. Essa atividade foi regida pelo operador *Reduce*, que somou os atributos masculinos, femininos, PS, S, MS, PSMasculinos, SMasculinos, MSMasculinos, PSFemininos, SFemininos e MSFemininos. Essa soma ocorreu totalizando, em uma única tupla do *DataFrame*, os dados resumidos.

Ao término de execução da atividade **Reduz_Estatisticas_Generos**, o resultado foi o *DataFrame* denominado **DF_Estatisticas_Generos**, o qual continha n=1 resumo estatístico dos PCU estabelecidos. Assim, a atividade **Reduz_Estatisticas_Generos** foi representada por **Reduce**(masculinos, femininos, PS, S, MS, PSMasculinos, SMasculinos, MSMasculinos, PSFemininos, SFemininos, MSFemininos). A Figura 38 ilustra o *schema* de **DF_Estatisticas_Generos** e a sua quantidade de tuplas.

O resultado da atividade **Reduz_Estatisticas_Generos** foi consumido imediatamente como entrada para a décima e última operação da álgebra de *Workflow*, denominada **Calcula_Totalizacoes**. Essa atividade foi regida pelo operador *Map*, que criou novos atributos, os quais foram usuarios, porcMasculinos, porcFemininos, porcPS, porcS e porcMS. Esses atributos foram criados para identificar as respectivas porcentagens masculinas e femininas dos PCU estabelecidos.

```
scala> DF_Estatisticas_Generos.printSchema()
root
|-- usuarios: long
|-- masculinos: long
|-- femininos: long
|-- PS: long
|-- S: long
|-- MS: long
|-- PSMasculinos: long
|-- SMasculinos: long
|-- MSMasculinos: long
|-- PSFemininos: long
|-- SFemininos: long
|-- MSFemininos: long

scala> DF_Estatisticas_Generos.count()
res50: Long = 1
```

Figura 38 – *Schema* e quantidade de tuplas de **DF_Estatisticas_Generos**

Ao término de execução da atividade **Calcula_Totalizacoes**, o resultado foi o *DataFrame* denominado **DF_Totalizacoes**, o qual continha n=1 resumo estatístico por gênero dos PCU estabelecidos. Assim, a atividade **Calcula_Totalizacoes** foi representada por **Map**(usuarios, porcMasculinos, porcFemininos, porcPS, porcS, porcMS). A Figura 39 ilustra o *schema* de **DF_Totalizacoes** e a sua quantidade de tuplas.

```
scala> DF_Totalizacoes.printSchema()
root
|-- usuarios: long
|-- masculinos: long
|-- porcMasculinos: double
|-- femininos: long
|-- porcFemininos: double
|-- PS: long
|-- porcPS: double
|-- S: long
|-- porcS: double
|-- MS: long
|-- porcMS: double
|-- PSMasculinos: long
|-- SMasculinos: long
|-- MSMasculinos: long
|-- PSFemininos: long
|-- SFemininos: long
|-- MSFemininos: long

scala> DF_Totalizacoes.count()
res52: Long = 1
```

Figura 39 – *Schema* e quantidade de tuplas de **DF_Totalizacoes**

4.2 Resultados das Análises Comportamentais com a identificação do PCU

Os resultados desta subseção foram produzidos pelo arcabouço de Análise do Padrão Comportamental dos Usuários. A população de usuários consistiu em 296 usuários, os quais usaram o arcabouço de Coleta dos Dados e tinham atividades físicas compartilhadas em seus perfis, as quais foram coletadas.

A Figura 40(a) ilustra a distribuição do gênero dos usuários. Houve a prevalência de usuários masculinos, em torno de 76%. A Figura 40(b) ilustra os PCU desta população. A maior parte dos usuários (59%) foi classificada com o PCU de Pouco Saudável. Mais adiante nessa subseção do texto, são discutidas algumas situações especiais onde os usuários foram classificados com o PCU de Pouco Saudável.

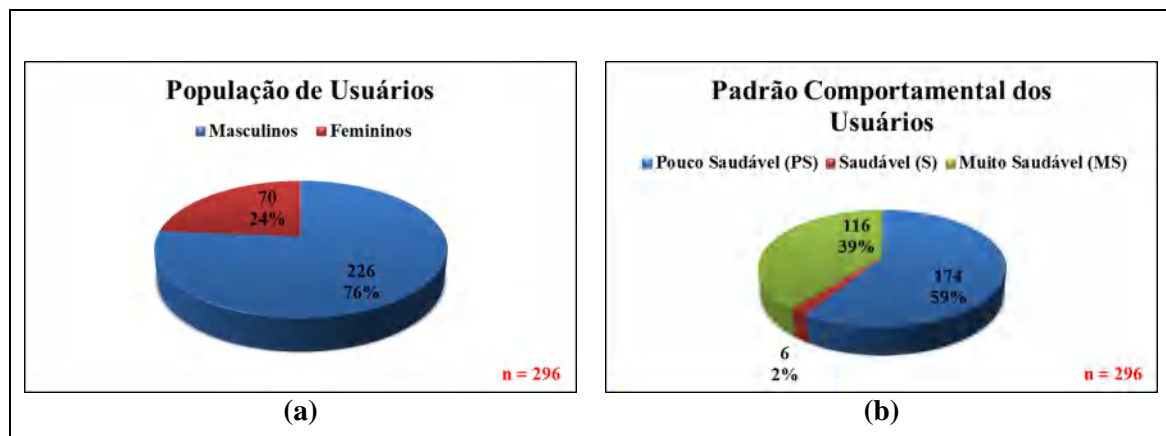


Figura 40 – População de usuários e PCU

A Figura 41(a) traz o comparativo dos PCU conforme o gênero do usuário. Houve uma aproximação dos valores dos PCU Pouco Saudável e Muito Saudável nos usuários masculinos. Na população estudada, 45% dos usuários masculinos foram classificados com os PCU Saudável ou Muito Saudável, bem como 30% dos usuários femininos foram classificados com os PCU Saudável ou Muito Saudável. Nenhum dos gêneros atingiu uma porcentagem de ao menos 50% com PCU Saudável ou Muito Saudável.

A Figura 41(b) ilustra a quantidade de semanas ativas desta população. A quantidade total de semanas ativas da população foi de 10.657. Houve a prevalência das semanas ativas masculinas, em torno de 81%. Em média, cada indivíduo masculino teve 38 semanas ativas, enquanto cada indivíduo feminino teve 28 semanas ativas. Esse

resultado sugestiona que os indivíduos masculinos foram mais ativos que os indivíduos femininos e, talvez, possa ser inferido que os indivíduos masculinos brasileiros tendem a ser mais ativos que os indivíduos femininos brasileiros, em proporções próximas dos resultados desta tese.

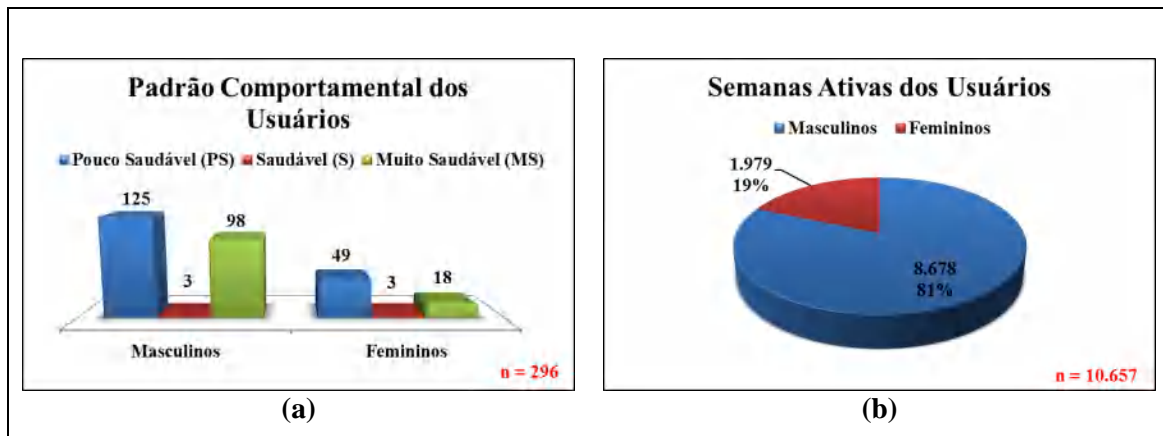


Figura 41 – PCU por gêneros e Semanas Ativas dos Usuários

A Figura 42(a) ilustra os PCU das semanas ativas. A maioria dos PCU semanais foi de Muito Saudável (MS), em torno de 50%, o que contrastou com o resultado dos PCU Finais desta população. Isso ocorreu pelo motivo de que os usuários com o PCU Muito Saudável (MS) tiveram mais semanas ativas do que os usuários com o PCU Pouco Saudável (PS).

A Figura 42(b) ilustra os PCU das semanas ativas conforme o gênero do usuário. Houve a prevalência dos PCU semanais Muito Saudável (MS) masculinos. Assim, 52% das semanas ativas foram Muito Saudáveis (MS) para os indivíduos masculinos e 41% das semanas ativas foram Muito Saudáveis (MS) para os indivíduos femininos.

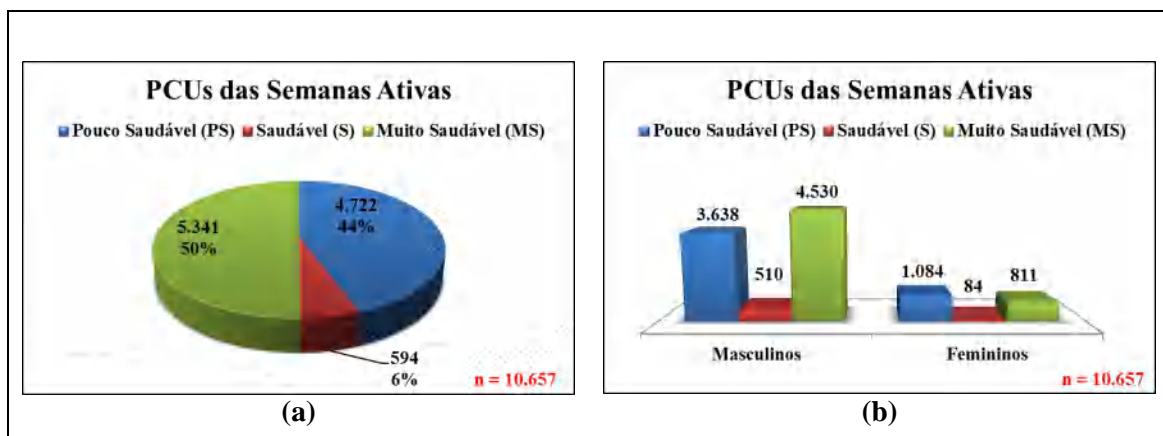


Figura 42 – PCUs das Semanas Ativas dos Usuários

A seguir, são apresentadas algumas análises complementares de comportamento. A Figura 43 ilustra a classificação adicional que foi inferida aos usuários da poluição. Foram criadas 15 classes esportivas e cada usuário da população foi associado a uma dessas classes esportivas. Isso foi feito de acordo com a frequência da prática das modalidades esportivas. Caso o usuário tenha praticado mais de uma modalidade ele foi associado pela ordem de frequência das modalidades. Um usuário associado à classe CiclistaCorredorCaminhante teve o ciclismo como modalidade mais frequente, a corrida como modalidade de frequência intermediária e a caminhada como modalidade de menor frequência.

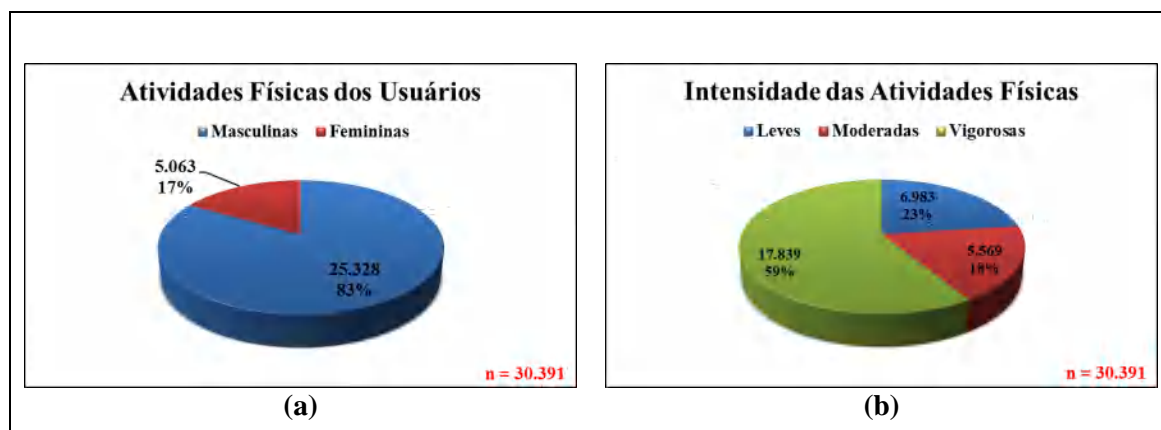


Figura 43 – Classes Esportivas dos Usuários

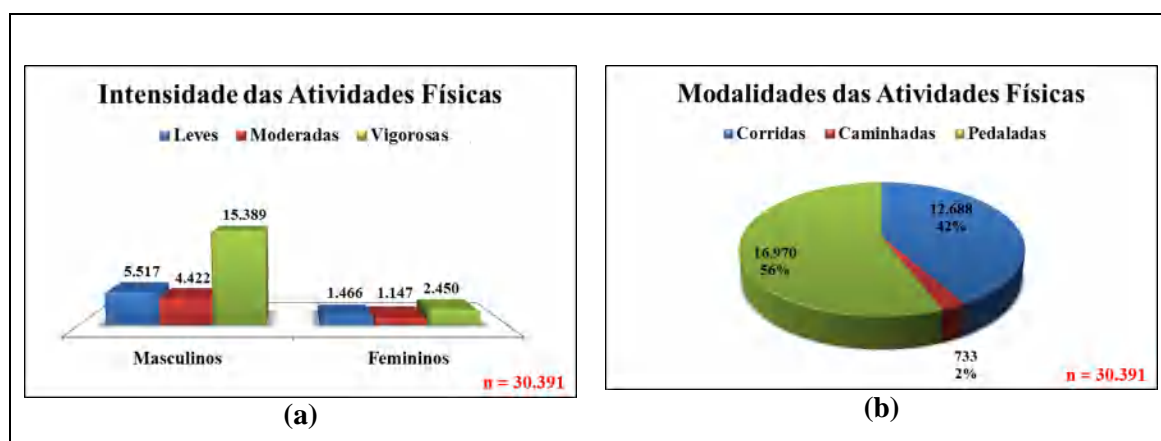
Houve a prevalência da classe de Corredores, seguidas das classes que combinaram corrida e ciclismo, e depois pela classe de Ciclistas. As classes que envolveram a modalidade de Caminhada, em sua maioria, tiveram uma baixa associação ou uma associação nula. Apesar da modalidade de Caminhada ser simples de ser executada, de poder ser executada com facilidade e em qualquer lugar, ela teve uma baixa frequência de associação nesta população. Isso pode sugerir que essa população teve o foco em praticar atividades mais intensas do que uma Caminhada.

A Figura 44(a) ilustra o total de atividades físicas praticadas pela população, conforme o gênero. Essa população praticou n=30.391 atividades físicas. Houve a prevalência das atividades físicas masculinas. Examinando os percentuais em relação a ambos os gêneros, ocorreram em média 112 atividades por indivíduos masculinos e 72 atividades por indivíduos femininos, correspondendo a 61% e 39%, respectivamente, do total de atividades médias por indivíduo para cada gênero.

A Figura 44(b) ilustra a intensidade dessas atividades físicas. Houve a prevalência das atividades físicas de intensidade vigorosa, em torno de 59%.



A Figura 45(a) ilustra intensidade dessas atividades físicas, conforme o gênero. Houve a prevalência das atividades físicas de intensidade vigorosa em ambos os gêneros. A Figura 45(b) ilustra as modalidades dessas atividades físicas. Houve a prevalência da modalidade de Pedaladas, em torno de 56%. Isso contrastou com o resultado das Classes Esportivas dos Usuários, já que a maioria dos usuários desta população é formada por Corredores. Isso sugere que, nesta população, os Ciclistas têm uma frequência de execução de atividades de ciclismo maior do que a frequência de execução de atividades de corridas pelos Corredores. Foi notada também a baixa frequência de execução da modalidade de Caminhadas, nesta população, em torno de 2%.



A Figura 46(a) ilustra as modalidades dessas atividades físicas, conforme o gênero. Houve a prevalência das atividades físicas masculinas em Pedaladas, mas houve

também a prevalência das atividades físicas femininas em Corridas. A Figura 46(b) ilustra as velocidades médias usuais em cada modalidade, conforme o gênero. De certa forma, as velocidades médias foram próximas nas mesmas modalidades, onde não ocorreram grandes variações entre os gêneros, porém com os valores das corridas e pedaladas masculinas ligeiramente superiores aos valores das corridas e pedaladas femininas.

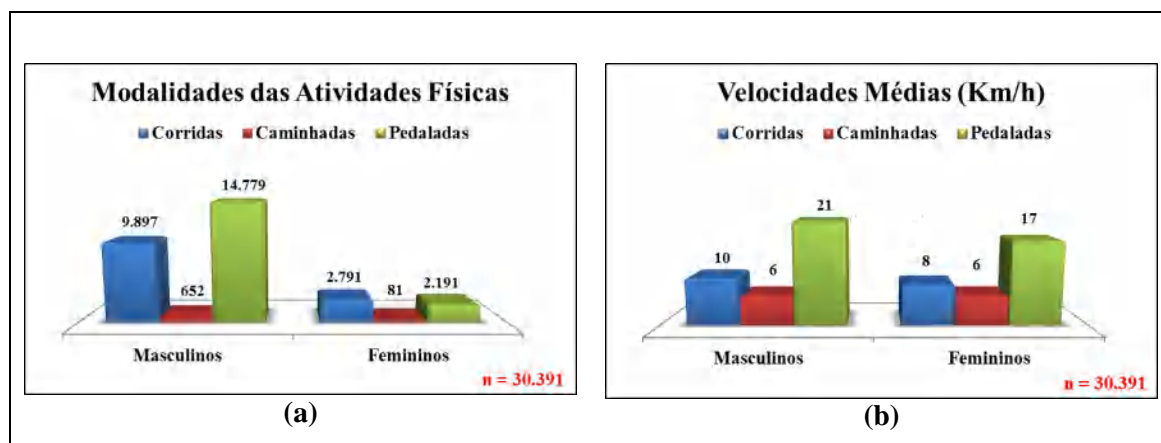


Figura 46 – Modalidades e Velocidades Médias

A Figura 47(a) ilustra o total de quilômetros percorridos nas modalidades dessas atividades físicas, conforme o gênero. Houve a supremacia das distâncias percorridas nas atividades físicas masculinas de Pedaladas, mas houve também a prevalência das distâncias percorridas nas atividades físicas femininas de Pedaladas. A Figura 47(b) ilustra o total de horas utilizadas na realização das atividades em cada modalidade, conforme o gênero. De forma análoga aos resultados das distâncias percorridas, houve a supremacia das durações nas atividades físicas masculinas de Pedaladas, mas houve também a prevalência das durações nas atividades físicas femininas de Pedaladas.

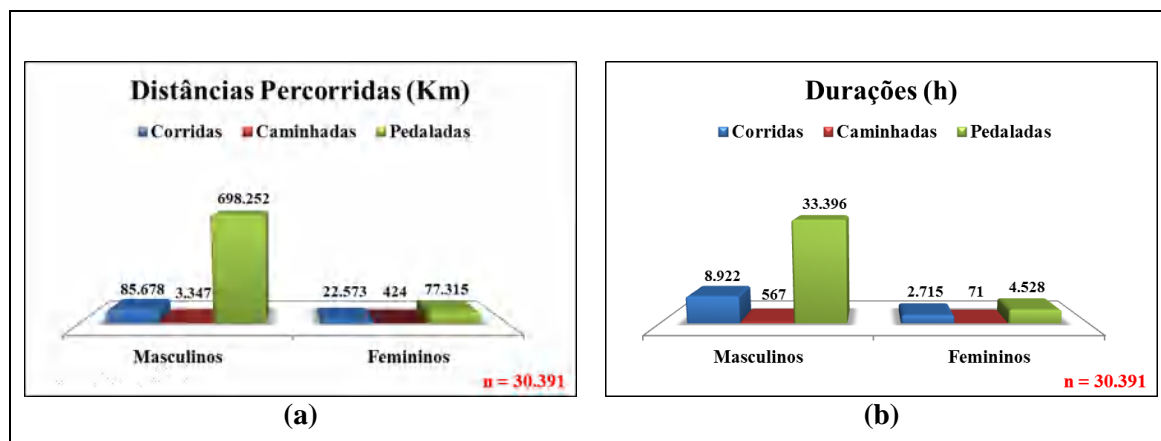


Figura 47 – Distâncias Percorridas e Durações

Em resumo, foi sugestionado o comportamento padrão do usuário desta população com base nos itens mais frequentes. Dessa forma, ele pode ter sido do gênero masculino, com o PCU Pouco Saudável (PS), mas com uma boa frequência de PCU semanais de Muito Saudável (MS). Ele pode ter sido corredor ou ciclista, com a maioria de suas atividades tendo intensidade vigorosa e a modalidade com maior frequência sendo o ciclismo.

Em relação aos gêneros, foi sugestionado o comportamento padrão desta população, com base nos itens mais frequentes. Assim, os usuários masculinos tiveram uma maior quantidade de semanas ativas que os usuários femininos, ambos os gêneros tiveram o PCU Pouco Saudável (PS), os usuários masculinos tiveram uma maior frequência de PCU semanais de Muito Saudável (MS) e os usuários femininos tiveram uma maior frequência de PCU semanais de Pouco Saudável (PS). Os usuários masculinos praticaram mais atividades físicas do que os usuários femininos e a maioria das atividades foram vigorosas para ambos os gêneros. A modalidade com maior frequência foi o ciclismo para os usuários masculinos e a corrida para os usuários femininos. Ambos os gêneros tiveram a modalidade de ciclismo com as maiores distâncias percorridas e também com as maiores durações de execuções das atividades.

A classificação dos PCU dos usuários foi feita com base no modelo sugestionado pela WHO e já detalhado na seção “3.5 Modelo de classificação de comportamento” desta tese. Com base nesse modelo, o PCU Saudável (S) de uma semana de atividades físicas praticadas foi inferido ao usuário que teve, no mínimo, 150 minutos de atividades físicas moderadas ou 75 minutos de atividades físicas vigorosas, na respectiva semana.

Nesse sentido, foram detectados casos com situações de exageros ou excessos da prática de atividades físicas, por alguns usuários da população estudada. Nesses casos, as semanas desses usuários com essas atividades exageradas ou excessivas foram classificadas com o PCU Pouco Saudável (PS), o que refletiu a sugestão de que os exageros ou excessos não representam algo saudável.

Algumas dessas situações foram percebidas como a prática de atividades com intensidades “explosivas”. Um exemplo disso foram as práticas de corridas “explosivas” com curtas durações. Então, um usuário que praticou 3 corridas semanais, com a distância percorrida de 5 km em cada uma dessas corridas, com duração inferior a 25 minutos em cada uma dessas corridas, teve a sua semana classificada como Pouco Saudável (PS).

Isso ocorreu pelo motivo de que a somatória das durações dessas 3 corridas dessa semana, que são atividades vigorosas mais intensas, não chegou ao limite mínimo de 75 minutos de atividades físicas vigorosas. De certa forma, essa classificação sugeriu um comportamento aparentemente correto, já que atividades explosivas podem aumentar as chances de o usuário ter alguma lesão ou problema de saúde pelo excesso de esforço físico. Assim, atividades que podem provocar lesões ou problemas de saúde sugerem um comportamento Pouco Saudável (PS).

Além das atividades com intensidades “explosivas”, foram detectados casos com atividades físicas com durações muito prolongadas e de intensidades “Leve”. Um exemplo disso foram as práticas de pedaladas de longas durações e de baixas velocidades médias. Então, um usuário que fez uma pedalada na semana, com a distância percorrida de 80 km, com duração de 8 horas, com velocidade média de 10 Km/h, teve a sua semana classificada como Pouco Saudável (PS).

Isso ocorreu pelo motivo de que essa pedalada foi classificada como uma pedalada de intensidade leve e não há previsão, no modelo de comportamento saudável sugerido pela WHO, da contabilização de atividades dessa intensidade. De certa forma, essa classificação sugeriu um comportamento aparentemente correto, já que uma pedalada a 10 km/h pode ser sugerida como uma velocidade baixa para uma atividade de ciclismo e, provavelmente, o esforço físico dessa atividade pode ser baixo. Assim, atividades de intensidade leve sugerem um comportamento Pouco Saudável (PS).

4.3 Avaliação de desempenho no processamento paralelo e distribuído

Essa seção apresenta a avaliação da execução do *Workflow* de modo paralelo e distribuído. A primeira etapa consistiu em produzir réplicas dos dados para ter um volume adequado para avaliação. A segunda e última etapa consistiu em avaliar o *Speedup*, a eficiência e o tempo de execução.

4.3.1 Volume de Dados

O arcabouço de Coleta de Dados iniciou o seu funcionamento há 8 meses, de forma estável e com acesso público. Ele tem sido utilizado há mais tempo por alguns usuários

que testavam as suas funcionalidades. Nesse curto período de tempo, a quantidade de usuários chegou a 378 usuários, abrangendo todos os usuários que acessaram esse arcabouço.

Desses 378 usuários, somente 296 usuários (78%) tinham atividades físicas compartilhadas em seus perfis por AS de monitoramento de atividades físicas. Dentre os 82 usuários (22%) que não tinham atividades físicas compartilhadas, a maior parte deles era formada por esportistas que não usavam AS de monitoramento de atividades físicas ou que usavam, mas não compartilhavam as suas atividades em seus perfis. Assim, os 296 usuários produziram o total de 30.391 atividades físicas compartilhadas. Apesar de ser um número expressivo, ele é relativamente baixo em comparação com os números envolvidos em análises e classificações de dados massivos (*BigData*) comumente tratados nos estudos de RSO.

Para o arcabouço de Coleta de Dados coletar um volume de dados equivalente ao de um problema de *BigData*, seria necessária a utilização do mesmo por um período de tempo bem maior que os 8 meses atuais, além de aumentar a forma de divulgação do mesmo para motivar a sua utilização. Esse volume relativamente baixo se deu pelo motivo do público alvo do arcabouço de Coleta de Dados ter sido bem seleto, já que eram usuários do Facebook que praticaram e compartilharam atividades físicas com os seus AS de monitoramento de atividades físicas e que também usaram o AS de Coleta de Dados.

Outro fator que ocorreu foi que alguns dos principais AS de monitoramento de atividades físicas modificaram a sua forma padrão de compartilhar as atividades físicas nos perfis do Facebook. O formato correto de compartilhamento é o formato de atividade física. Apesar de eles continuarem mantendo esse formato disponível para o usuário, o qual passou a requerer uma pequena configuração adicional no AS de monitoramento de atividades físicas, a forma padrão de compartilhamento mudou para o padrão de foto.

Nesses casos com o compartilhamento no formato de foto, não foi possível coletar os dados. Os usuários do arcabouço de Coleta de Dados que tiveram esse problema foram auxiliados no ajuste dessa configuração adicional, em seu AS de monitoramento de atividades físicas, os quais passaram a compartilhar no formato correto. Assim, na maioria dos casos esse problema foi contornado, mas com a perda da coleta do compartilhamento de algumas atividades físicas postadas no formato incorreto.

Somente os usuários das versões mais recentes do AS Nike + Running, denominado Nike + Club, não tiveram esse problema contornado pelo fato de não estar mais disponível o compartilhamento no formato de atividade física. Nesses casos, os usuários foram orientados a instalar a versão anterior do AS Nike + Running e não atualizá-lo para a nova versão.

Dessa forma, os resultados obtidos por essa tese foram feitos com base na população de, no máximo, 378 usuários. Alguns subgrupos dessa população foram formados para algum estudo específico, como os 296 usuários que tinham atividades físicas compartilhadas e que foram estabelecidos os seus PCU.

Porém, para efeito de avaliação de desempenho do arcabouço de Análise do Padrão Comportamental dos Usuários, com o processamento paralelo e distribuído, analisando um grande volume de dados, foi produzida, de forma adicional ao *Framework* desta tese, uma replicação dos dados. Essa replicação não foi considerada no *Framework* e nos resultados da tese, tendo sido considerada apenas para o estudo do *Speedup* da execução deste arcabouço. Isso se deu pelo fato de não ter havido tempo hábil para a formação natural do volume de dados necessário, o que configuraria um *BigData*, até a conclusão desta tese.

4.3.2 Análise de resultados

A avaliação do desempenho foi feita com base em um conjunto de arquivos replicados, a partir do arquivo original obtido pelo arcabouço de Coleta dos Dados. Essa replicação foi feita para avaliar o desempenho do *Workflow* no processamento paralelo e distribuído com um volume de dados equivalente a um *BigData*, a qual verificou a escalabilidade do *Workflow*. Inicialmente, será explicada nessa seção como foi feita a replicação dos dados. Posteriormente, serão apresentados os resultados da avaliação de desempenho, como o *Speedup*, a Eficiência e o Tempo de Execução.

A replicação dos dados consistiu na criação de arquivos CSV contendo a primeira linha com os nomes dos atributos e as demais linhas com as atividades físicas compartilhadas pelos usuários do arcabouço de Coleta dos Dados. O arquivo original tinha 30.391 atividades físicas compartilhadas, com o tamanho de aproximadamente 4 *Mega Bytes* (MB). Assim, foram criados arquivos CSV com a replicação dessas 30.391

atividades, para o processamento paralelo e distribuído das réplicas, com o limite máximo de réplica contendo mais de 583 milhões de atividades físicas.

Os arquivos replicados contiveram múltiplos do conteúdo do arquivo original, para ocorrer a replicação de forma integralizada dos dados. Isso significou que não ocorreu a replicação de um conjunto parcial dos dados, ao final do processo de replicação. A Tabela 4 ilustra as características dessas réplicas. O Rótulo foi criado para identificar o arquivo no processamento paralelo e distribuído. O Tamanho (MB) indica o tamanho real do arquivo, em MB. A quantidade das Atividades Físicas é ilustrada em cada arquivo. O Multiplicador indica a quantidade de vezes que os dados originais foram replicados no respectivo arquivo. O arquivo de Rótulo 4MB foi o arquivo original que foi replicado.

Portanto, o arquivo original, de Rótulo 4MB, foi replicado em 13 arquivos maiores, com o fator multiplicador máximo de 19.200 na maior réplica, a qual foi representada pelo arquivo de Rótulo 76.800MB, contendo um total de 583.507.200 atividades físicas e que foram processadas, o que permitiu a verificação da escalabilidade do *Workflow*.

Assim, o *Speedup* do *Workflow* avaliou o desempenho de processamento paralelo e distribuído do arcabouço de Análise do Padrão Comportamental dos Usuários, com o processamento paralelo e distribuído desses 14 arquivos de tamanhos variados.

Tabela 4 – Replicação dos Dados

Replicação dos Dados			
Rótulo	Tamanho (MB)	Atividades Físicas	Multiplicador
4MB	4,6	30.391	1
40MB	45,5	303.910	10
80MB	91,0	607.820	20
400MB	455,2	3.039.100	100
800MB	910,4	6.078.200	200
1.600MB	1.820,7	12.156.400	400
2.400MB	2.731,1	18.234.600	600
3.200MB	3.641,5	24.312.800	800
4.000MB	4.551,8	30.391.000	1.000
4.800MB	5.462,2	36.469.200	1.200
9.600MB	10.924,4	72.938.400	2.400
19.200MB	21.848,7	145.876.800	4.800
38.400MB	43.697,4	291.753.600	9.600
76.800MB	87.394,8	583.507.200	19.200

Cada arquivo foi submetido ao processamento do Apache Spark, para cada configuração do número de núcleos utilizados no processamento. Na ocasião de

execução do *Speedup*, estavam disponíveis 4 núcleos em 3 nós. Assim, cada arquivo replicado foi processado pela primeira vez com apenas 1 núcleo (1 nó), pela segunda vez com 4 núcleos (1 nó), pela terceira vez com 8 núcleos (2 nós) e pela quarta vez com 12 núcleos (3 nós).

Dessa forma, o *Speedup* teve com base o tempo de processamento dos arquivos da Tabela 4. Além disso, para evitar uma interferência que possa ter ocorrido na ocasião de execução do *Speedup*, foram realizadas três rodadas de execuções, com cada rodada tendo sido feita em dias e horários diferentes. Então foi calculada a média dos tempos de execuções dessas três rodadas, representada na Tabela 5. Esses valores médios foram utilizados nos cálculos do *Speedup*.

Tabela 5 – Médias das execuções para o *Speedup*

Execuções no cluster (tempo em segundos)							
Índice	Arquivo	1 Processador	4 Processadores	8 Processadores	12 Processadores	Atividades Físicas	Bytes
		aldebaran	aldebaran	kaus aldebaran	kaus aldebaran rukbat		
1	4MB	133	61	72	81	30.391	4.773.111
2	40MB	119	60	47	39	303.910	47.729.464
3	80MB	128	62	47	40	607.820	95.458.744
4	400MB	217	82	59	47	3.039.100	477.292.984
5	800MB	328	111	75	58	6.078.200	954.585.784
6	1.600MB	552	159	104	99	12.156.400	1.909.171.384
7	2.400MB	754	220	132	118	18.234.600	2.863.756.833
8	3.200MB	958	262	164	133	24.312.800	3.818.342.433
9	4.000MB	1.185	296	183	146	30.391.000	4.772.928.033
10	4.800MB	1.401	337	219	173	36.469.200	5.727.513.784
11	9.600MB	2.731	651	366	302	72.938.400	11.455.027.384
12	19.200MB	5.291	1.235	889	649	145.876.800	22.910.054.584
13	38.400MB	11.482	2.427	1.872	1.503	291.753.600	45.820.108.984
14	76.800MB	21.075	4.438	3.510	3.287	583.507.200	91.640.217.784

A Figura 48 ilustra o cálculo de *Speedup* executado conforme a quantidade de processadores utilizados no processamento. Nela também foi apresentado o *Speedup* ideal, o qual é a medida desejada que o algoritmo obtenha em sua execução paralela, sendo proporcional a quantidade de processadores. Assim, o *Speedup* ideal tem o seu valor igual ao número de processadores utilizados. O *Speedup* pode ser classificado de três formas diferentes, de acordo com as características do algoritmo paralelizado, incluindo a sua complexidade e a sua necessidade de comunicação dos processadores.

Quando o *Speedup* é inferior ao *Speedup* ideal, ele é classificado como *Speedup* sublinear. Quando é igual ao *Speedup* ideal, ele é classificado como *Speedup* linear. Quando é superior ao *Speedup* ideal, ele é classificado como *Speedup* superlinear.

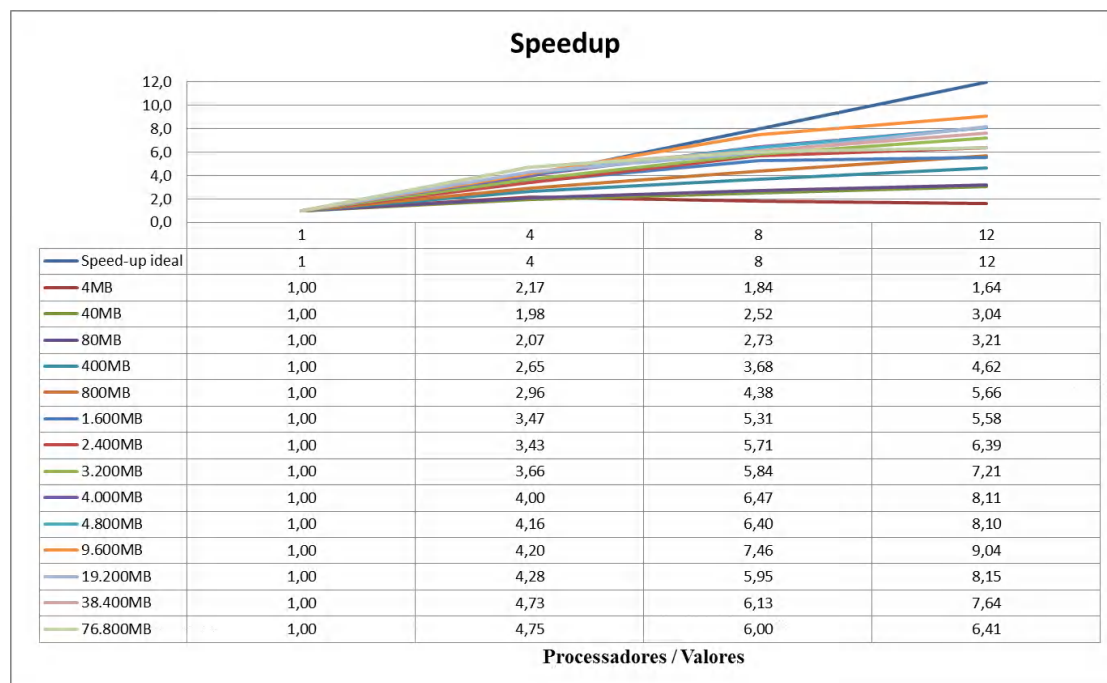


Figura 48 – Speedup de Execução

Assim, o *Speedup* desta tese foi classificado como *Speedup* sublinear. Isso ocorreu devido ao fato que os algoritmos complexos ou que precisem de muita comunicação entre os processadores normalmente têm seu *Speedup* classificado como sublinear. Outra observação que foi sugestionada pelo resultado foi que o *Speedup* foi melhorando conforme o aumento do volume de dados, onde os processamentos com os maiores arquivos proporcionaram os melhores resultados, no limite do tamanho de arquivo de 9.600MB.

O desempenho começa a cair a partir dos arquivos maiores que o tamanho limite de 9.600 MB. Isso se deu pelo motivo desses arquivos excederem o tamanho da memória disponível para processamento.

Em termos de eficiência do algoritmo, ele teve um desempenho razoável. O desempenho também teve o comportamento similar ao *Speedup*, ou seja, foi sugestionado que ela foi melhorando conforme o aumento do volume de dados, onde os processamentos com os maiores arquivos foram os mais eficientes, no limite do tamanho de arquivo de 9.600MB. A Figura 49 ilustra o cálculo de Eficiência do algoritmo conforme a quantidade de processadores utilizados em seu processamento.

Porém, mesmo o algoritmo não tendo tirando o máximo proveito do ambiente paralelizado, ele escalou bem e foi beneficiado por um tempo de execução bem menor em relação ao tempo de execução em um único processador.

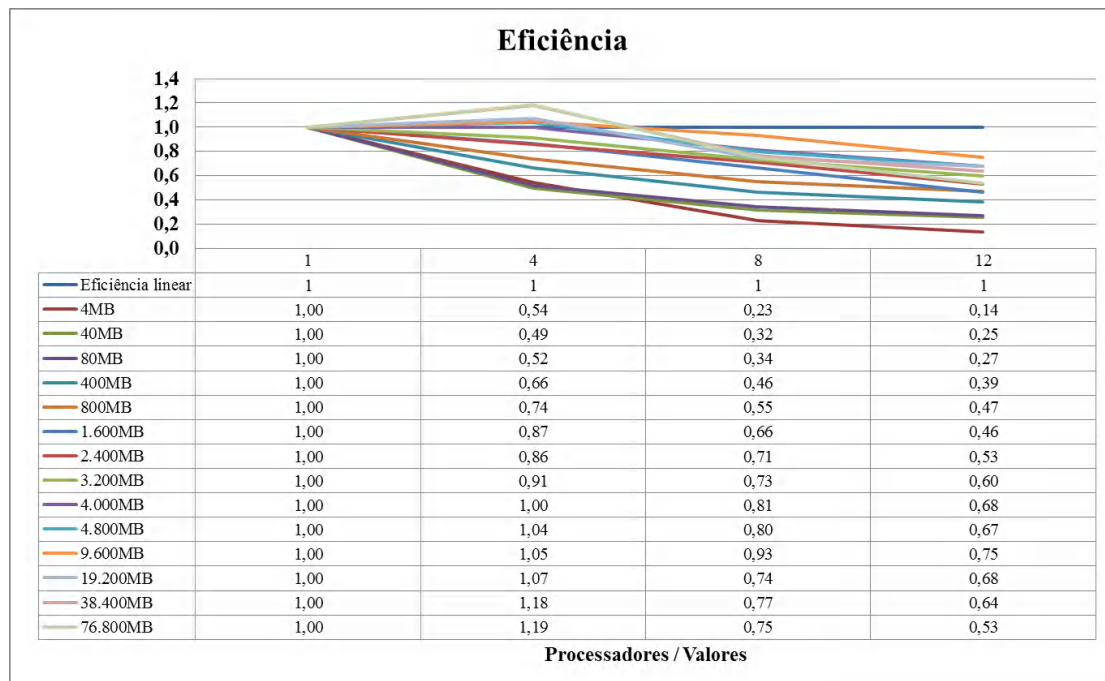


Figura 49 – Eficiência de execução

A Figura 50 ilustra o tempo de execução do algoritmo, em segundos, conforme a quantidade de processadores utilizados em seu processamento. Esses resultados sugestionam que o ambiente Apache Spark, em conjunto com o algoritmo desenvolvido, teve um desempenho inferior com baixo volume de dados e um melhor desempenho com o aumento significativo do volume de dados.

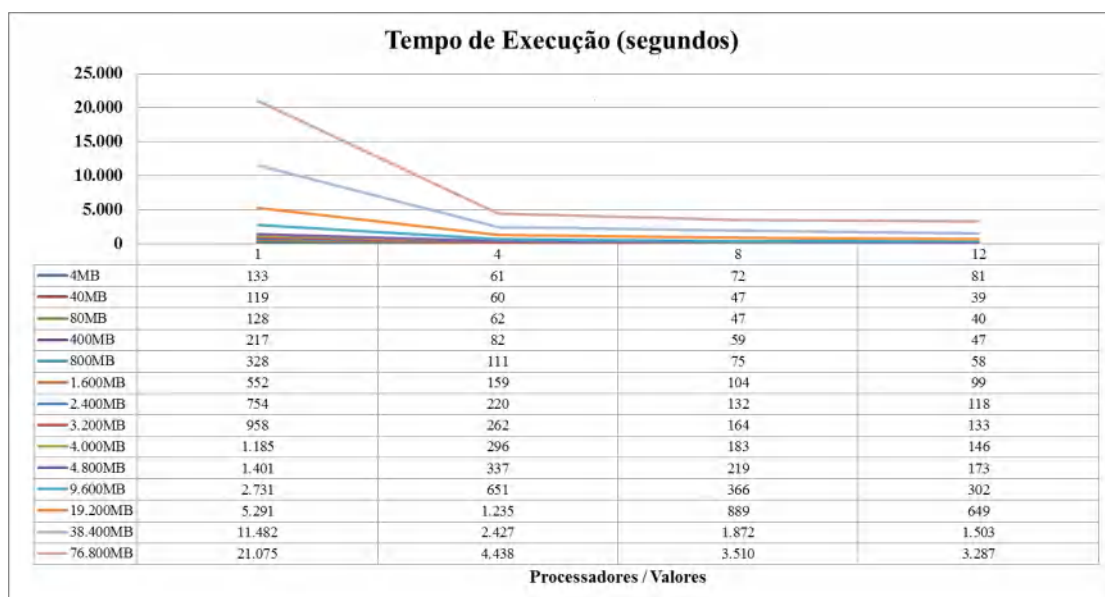


Figura 50 – Tempo de execução

4.4 Discussões

Em resumo, a avaliação experimental permitiu a avaliação da modelagem dos *Workflows* propostos, da identificação dos PCU e da execução do *Workflow* de modo paralelo e distribuído.

Dessa forma, esses resultados dessas avaliações corroboraram e responderam, de forma positiva, a Hipótese proposta por essa tese, ao menos no cenário de *fitness*, pois estabeleceu os Padrões Comportamentais dos Usuários (PCU):

A evolução do uso dos aplicativos sociais no Facebook pode identificar padrões comportamentais dos usuários no contexto de fitness.

Assim, os usuários tiveram os seus PCU definidos conforme as suas práticas de atividades físicas compartilhadas pelos AS de monitoramento de atividades físicas em seus perfis do Facebook.

De forma complementar, os resultados sugeriram a importância de se estabelecer o PCU de comportamento saudável, principalmente pelo motivo dos excessos da prática esportiva, os quais resultaram em um comportamento Pouco Saudável. Os AS de monitoramento de atividades físicas focam no monitoramento da atividade física e na disponibilização de *rankings* competitivos, mas não se preocupam disponibilizar ao usuário informações sobre se o comportamento é saudável ou não. Eles fazem isso, no máximo, com o monitoramento cardíaco indicando as faixas de esforço.

Então, muitos usuários que figuram nas primeiras posições desses *rankings* criados pelos AS podem ter um comportamento Pouco Saudável e não sabem disso, pois podem estar focados apenas em estar nas posições iniciais do *ranking*. Assim, é importante a disponibilização dos PCU, de forma que o usuário tenha uma prática esportiva consciente e saudável.

Um fato interessante ocorreu com a população estudada, com a predominância de homens casados mais jovens e mulheres solteiras mais maduras. Além disso, os usuários compartilham as suas atividades físicas há um tempo considerável, de pelo menos 2 anos, com uma quantidade razoável de atividades físicas compartilhadas, de pelo menos 100 atividades físicas compartilhadas. E os homens praticaram mais ciclismo e as mulheres praticaram mais corridas.

E em relação ao processamento paralelo e distribuído, o *Workflow* escalou bem e atingiu os objetivos de fornecer uma arquitetura paralela e distribuída de classificação dos usuários do Facebook.

Capítulo 5 - Trabalhos Relacionados

Esse capítulo apresenta os trabalhos relacionados ao assunto desta tese. Ele foi subdividido em três seções. A primeira seção apresentará algumas das análises comportamentais em redes sociais. A segunda seção apresentará alguns *Frameworks* que fazem a análise de comportamento em redes sociais. A terceira seção apresentará algumas ferramentas de análise de *fitness*. Nestas seções, a apresentação dos trabalhos é feita em ordem cronológica decrescente. Ao final, são apresentadas as considerações finais sobre os trabalhos relacionados.

5.1 Análise comportamental em Redes Sociais

Essa seção descreve as ferramentas de análise de comportamento dos usuários de redes sociais, mas sem o uso de aplicativos sociais para a análise, coleta ou a postagem de dados nos perfis dos usuários.

5.1.1 Predição de status de saúde com o uso de Apache Spark com *Machine Learning*

Nair et al. (2017) desenvolveram um sistema remoto de previsão de status de saúde, em tempo real, com o uso do processamento paralelo e distribuído oferecido pelo Apache Spark com análises de *Machine Learning*. Esse sistema trabalha com o volume dos dados de *BigData*, os quais são originados do Twitter. O sistema é escalável e analisa os dados da RSO Twitter, fazendo a predição do estado de saúde dos usuários (Nair et al. 2017).

A semelhança deste sistema com os arcabouços desenvolvidos nesta tese é a análise de uma RSO com o uso do paralelismo do Apache Spark, de forma escalável.

As diferenças entre esse estudo e os arcabouços desenvolvidos nesta tese compreendem que esse estudo utilizou uma RSO diferente (Twitter x Facebook), o foco de estudo é diferente, não utilizou uma competição social e nem definiu um PCU.

5.1.2 Influência das mídias sociais nos comportamentos saudáveis de jovens adultos

Vaterlaus et al. (2015) propuseram um estudo do uso do Facebook para motivar os jovens adultos a ter um comportamento mais saudável. Foi percebida que a exposição no Facebook, das atividades físicas compartilhadas pelos AS de monitoramento de atividades físicas, podem motivar os amigos desses usuários à prática de atividades físicas e também à mudança comportamental em busca de comportamentos mais saudáveis (Vaterlaus et al. 2015).

A percepção da mudança comportamental desse estudo reforça a importância dos arcabouços desenvolvidos nesta tese, no sentido desses arcabouços poderem motivar seus usuários a terem um estilo de vida mais saudável. Essa percepção foi a única semelhança entre esse estudo e os arcabouços desenvolvidos nesta tese.

As diferenças entre esse estudo e os arcabouços desenvolvidos nesta tese compreendem que esse estudo utilizou questionário para fazer a coleta dos dados, incluindo inclusive outra característica não utilizada no desenvolvimento desta tese, como hábitos alimentares. Ele não utilizou uma competição social e nem definiu um PCU.

5.1.3 Mídias sociais *online* para aumentar o nível de atividades físicas

Zhang et al. (2015) fizeram um estudo demonstrando a influência das mídias sociais *online* sobre o nível de atividades físicas dos usuários da RSO. Foi analisado o quanto a exposição nas mídias sociais *online* pode contribuir na prática de atividades físicas entre estudantes universitários. Os resultados demonstraram que os estudantes expostos praticaram mais atividades físicas do que os que não foram expostos (Zhang et al. 2015).

A semelhança entre esse estudo e os arcabouços desenvolvidos nesta tese está na sugestão de que a exposição das atividades físicas nas RSO estimulam a prática de atividades físicas, de forma a motivar o aumento dessa prática com o engajamento dos amigos de seus usuários.

As diferenças são que o estudo não desenvolveu um aplicativo social para promover a prática de atividades físicas, também não fez uma competição social entre

os participantes, não fez classificações da prática de atividades físicas e não definiu um PCU.

5.1.4 Intervenção de atividades físicas baseadas em mídias sociais

Cavallo et al. (2012) propuseram um estudo do uso do Facebook para motivar a prática de atividades físicas em estudantes universitários. Foram criados dois grupos aleatórios de estudantes, com ambos os grupos acessando um *website* com foco em atividades físicas, onde um desses grupos foi estimulado também a utilizar um grupo do Facebook, bem como a fazer atividades auto-monitoradas, mas sem o uso de aplicativos sociais (Cavallo et al. 2012).

Esse estudo concluiu que não houve diferenças significativas no nível da prática de atividades físicas dos dois grupos. Portanto, a esperada possibilidade de aumento da prática de atividade física, pelo uso do Facebook, não ocorreu.

Assim, a semelhança entre este estudo e os arcabouços desenvolvidos nesta tese é que ambos utilizaram o Facebook para tentar motivar a prática de atividades físicas. As diferenças estão no fato desse estudo não usou um aplicativo social, não fez a coleta dos dados no Facebook, não ofereceu uma competição social entre os seus usuários e não definiu um PCU.

5.1.5 Análise de conjuntos de dados de redes sociais

Xue et al. (2010) desenvolveram a biblioteca de código aberto X-RIME para a análise de conjuntos de dados massivos de redes sociais, baseada no modelo de programação *MapReduce*. Essa biblioteca pode ser integrada com ferramentas de análise de dados e de estatísticas, como JAQL e Mahout, para a construção de soluções mais compreensivas de *Business Intelligence* (BI) (Xue et al. 2010).

Essa biblioteca tem a sua arquitetura formada por algumas camadas. Dentre elas, há a camada de modelo de dados que utiliza o *MapReduce* do Hadoop. Foram desenvolvidos programas de *MapReduce* para remover as redundâncias nas ligações entre os nós da rede social (caso dois nós tenham mais de um relacionamento, esse relacionamento é reduzido para um único relacionamento) com o objeto de simplificar

esse relacionamento. No *cluster* computacional foram utilizados sete nós com dois processadores em cada nó (Xue et al. 2010).

A semelhança entre esse trabalho relacionado e os arcabouços desenvolvidos nesta tese é que ambos trabalharam com *MapReduce* na análise dos dados dos usuários do Facebook, e usaram um sistema paralelo e distribuído para processar as análises.

As diferenças estão no fato desse trabalho relacionado não usou um aplicativo social, não ofereceu uma competição social entre os seus usuários e não definiu um PCU.

5.1.6 Bem estar em redes sociais

Burke et al. (2010) fizeram um estudo de como os usuários de redes sociais, como o Facebook, se comportam e quais são os benefícios sociais que o uso dessas redes traz aos seus usuários (Burke et al. 2010).

Foram classificadas duas formas distintas de atividades sociais nessa rede. A primeira forma ocorre quando o usuário se relaciona e interage ativamente com um grupo de pessoas próximas a ele e, nesse caso, é formada por um pequeno grupo de pessoas. A segunda forma ocorre quando o usuário acompanha as atualizações públicas de seus amigos, como atualizações de *status* ou conversas públicas de seus amigos com outros amigos (que muitas vezes ele não conhece) e, nesse caso, o usuário utiliza a rede social de forma passiva (Burke et al. 2010).

Os dados das atividades de um grupo de usuários do Facebook foram coletados durante dois meses, os quais foram agregados à plataforma computacional distribuída do Hadoop. Os dados foram coletados com o uso de questionários e também por *logs* das atividades realizadas pelos usuários no Facebook, com o processamento paralelo e distribuído tendo feito a análise dos *logs*. Com isso, foi verificado como esses dados estão correlacionados (Burke et al. 2010).

A semelhança entre esse trabalho relacionado e os arcabouços desenvolvidos nesta tese é que ambos estudaram algum tipo de comportamento do usuário no Facebook, bem como foram processadas as análises em um sistema paralelo e distribuído.

As diferenças estão no fato desse trabalho relacionado não usou um aplicativo social, não ofereceu uma competição social entre os seus usuários e não definiu um

PCU. Além disso, a coleta dos dados dos arcabouços desenvolvidos por esta tese criaram panoramas de análises dos dados para definir o comportamento do usuário, o que não foi feito nesse trabalho relacionado.

5.2 *Framework* para análise de comportamento em Redes Sociais

Essa seção descreve os aplicativos sociais de análise de comportamento de usuários de redes sociais, com o uso de aplicativos sociais para a análise, coleta ou a postagem de dados nos perfis dos usuários.

5.2.1 Usuários Novatos x Veteranos de AS de monitoramento de atividades físicas

Stragier et al. (2016) fizeram um estudo comparativo com os usuários esportistas de AS de monitoramento de atividades físicas, como o Strava, junto com tecnologias de vestir, onde levantaram algumas motivações de uso do Strava ou tecnologias de vestir, por parte dos usuários novatos e também dos usuários veteranos. Foram estudados os motivos dos usuários não usarem por muito tempo o Strava ou tecnologias de vestir e, principalmente, quais foram os fatores que mantiveram esses usuários em uma utilização por um período prolongado (Stragier et al. 2016).

Foi desenvolvido um AS para a coleta dos dados das atividades físicas compartilhadas no Twitter, feitas com o uso do Strava. Com base nos dados coletados no Twitter, foram extraídas informações complementares no site do Strava, através da URL disponibilizada no compartilhamento feito no Twitter e com o uso de um *script* PHP para esse fim. Esse *script* identificou os usuários que finalizaram com sucesso a atividade física compartilhada no Twitter. Finalmente, foi enviado um questionário eletrônico para esses usuários, para obter as informações a serem analisadas.

O estudo concluiu que tanto os AS de monitoramento de atividades físicas quanto as tecnologias de vestir deveriam oferecer uma experiência social agradável, incluindo a disponibilização dos dados monitorados. Assim, facilitar ao usuário o compartilhamento dessas informações nas RSO é um fator decisivo no sucesso de uso por um período prolongado.

A semelhança entre esse estudo e os arcabouços desenvolvidos nesta tese compreende o desenvolvimento de um AS de Coleta dos Dados de postagem criadas por AS de monitoramento de atividades físicas.

As diferenças entre esse estudo e os arcabouços desenvolvidos nesta tese estão na RSO de estudo (Twitter x Facebook), na coleta dos dados compartilhados (Strava x 16 AS de monitoramento de atividades físicas), no pré-processamento dos dados a serem analisados (questionário eletrônico x AS de Coleta dos Dados), dentre outras diferenças. Nesse estudo não foi utilizado o processamento paralelo e distribuído, bem como não foi feita uma competição social entre os usuários e não definiu um PCU.

5.2.2 Motivação de atividades físicas no trabalho

Foster et al. (2010) fizeram um estudo onde sugeriram a importância de uma competição social e interativa para motivar a mudança comportamental para um estilo de vida mais saudável. Foram estudados dois grupos distintos, com o uso de um AS desenvolvido para a coleta do número de passos realizados no ambiente de trabalho, bem como para disponibilizar ou não as informações das passadas aos integrantes dos grupos. O estudo foi composto por 10 colegas de trabalho (9 mulheres e 1 homem), com o tempo de uso do Facebook de no mínimo 1 ano e com ambos os integrantes do grupo sendo amigos no Facebook. No primeiro grupo (5 pessoas), cada membro só teve acesso aos dados de suas passadas. Já no segundo grupo (5 pessoas), todos os membros tiveram acesso aos dados das passadas de todos os membros, através da disponibilização de um *ranking* competitivo e do compartilhamento da execução das passadas. Foi notado no segundo grupo, que teve o apoio da competição social, que houve um aumento significativo do número de passos. Isso sugeriu que o uso de AS competitivos pode desempenhar um papel importante no estímulo da prática de atividades físicas, de uma forma motivadora e divertida (Foster et al. 2010).

A quantidade de passos realizada por cada integrante foi lançada manualmente no AS desenvolvido. No grupo com apoio da competição social, o AS notificava automaticamente os outros membros do grupo sobre a ocorrência dessa atividade física, bem como compartilhava a mesma no perfil do usuário. Assim, essa atividade física era passível de comentários entre os membros do grupo.

As semelhanças entre este estudo e os arcabouços desenvolvidos nesta tese compreendem o desenvolvimento de um AS para a Coleta dos Dados, com o seu uso na mesma RSO (Facebook), a criação de uma competição social saudável entre os seus usuários e de serem mecanismos de motivação e de promoção à mudança de estilo de vida saudável.

As diferenças entre este estudo e os arcabouços desenvolvidos nesta tese compreendem a forma de coleta dos dados (entrada manual x entrada automatizada), a origem dos dados (pedômetro x compartilhamentos dos AS de monitoramento de atividades físicas), a variedade de tipos de atividades físicas (passadas x caminhadas, corridas e pedaladas), a possibilidade de variações do *ranking* competitivo (0 variações x 60 variações), a definição do PCU saudável do usuário não foi abordada no estudo e esse estudo não foi proposto de forma a ser utilizado o processamento paralelo e distribuído.

5.3 Ferramentas de análise de *fitness*

Essa seção descreve as ferramentas de análise de *fitness* dos usuários de redes sociais, mas sem o uso de aplicativos sociais para a análise, coleta ou a postagem de dados nos perfis dos usuários. Essas ferramentas de análise de *fitness* não funcionam de forma integrada ao Facebook.

5.3.1 MovaMais

O MovaMais é um aplicativo que coleta os dados de alguns aplicativos sociais de monitoramento de atividades físicas, com o Strava, RunKeeper e MapMyRun. Ele cria uma espécie de bonificação aos seus usuários, como um programa de recompensas, conforme a prática regular de atividades físicas desses usuários (Aquino e Gomes 2016).

As atividades físicas que são convertidas em bonificações compreendem caminhadas, corridas e ciclismo. Essas bonificações podem ser trocadas por pontos ou benefícios de outros programas de recompensas brasileiros, como pontos Multiplus para o resgate de passagens aéreas ou para compras de produtos (Aquino e Gomes 2016).

As semelhanças entre o MovaMais e os arcabouços desenvolvidos nesta teste compreendem a motivação ao usuário para passar a ter um estilo de vida com hábitos mais saudáveis da prática regular de atividades físicas; bem como ambos utilizam os dados produzidos pelos AS de monitoramento de atividades físicas em *smartphones*.

As principais diferenças entre o MovaMais e os arcabouços desta tese compreendem: o MovaMais coleta as informações diretamente dos AS de monitoramento de atividades físicas no *smartphone*, sem que seja necessário o compartilhamento da atividade no Facebook, enquanto os arcabouços desta tese fazem a coleta dos perfis do Facebook; o MovaMais é um instrumento de motivação individual do usuário à prática de atividades físicas, com o apelo financeiro, enquanto os arcabouços desta tese têm a motivação coletiva com a sociabilidade do Facebook e com a competição social de seus usuários motivando a prática de atividades físicas. E o MovaMais não definiu um PCU.

5.3.2 Escolha de amigo para a prática de atividade física

Essa ferramenta permite a seleção de amigos para a prática de atividades físicas em grupo (Centola e Rijt 2015). Ele utiliza uma competição de *fitness* e um sistema de *ranking*, onde permite a seleção de amigos conforme as similaridades ou aproximações das características de execução de atividades físicas, como gênero, idade e preferências de execuções de atividades físicas, dentre outras características.

Ela utiliza as características das mídias sociais. As preferências de execuções de atividades físicas compreendem a intensidade de execução, a duração da atividade física, as modalidades de atividades físicas e a quantidade de atividades físicas.

Os resultados demonstraram que os usuários selecionaram os seus amigos de forma similar à forma de seleção *off-line*, levando em consideração as características de gênero, idade e IMC, não explorando toda a potencialidade de seleção de amigos conforme a aptidão física.

As semelhanças entre essa ferramenta e os arcabouços desenvolvidos nesta tese limitam-se à motivação da prática esportiva em grupos, com a disponibilização de diferentes informações da prática de atividades físicas entre os amigos. Além disso, há a classificação dos usuários conforme as características de aptidão física.

As diferenças estão no fato que essa ferramenta busca fornecer um amigo à execução de atividades físicas, enquanto os arcabouços buscam criar uma competição social com a classificação do usuário e de seus amigos. Essa ferramenta não trabalha de forma integrada ao Facebook, já os arcabouços trabalham. E essa ferramenta não definiu um PCU.

5.3.3 Saúde baseada na Internet

Lu et al. (2014) fizeram um estudo demonstrando a importância da criação de *sites* de Internet que promovam a prática de atividades físicas, de forma que utilizem um bom nível de interatividade e com a riqueza nos conteúdos das mídias. Foi analisado o quanto esses fatores ajudaram a promover a prática de atividades físicas entre estudantes universitários. Os resultados demonstraram que esses fatores auxiliaram, de forma positiva, o engajamento desses estudantes na prática de atividades físicas (Lu et al. 2014).

A semelhança entre esse estudo e os arcabouços desenvolvidos nesta tese está no fato da preocupação em disponibilizar uma riqueza de conteúdo da prática de atividades físicas, de forma a motivar o aumento dessa prática com o engajamento de seus usuários.

As diferenças são que o estudo não desenvolveu um aplicativo social para promover a prática de atividades físicas, também não fez uma competição social entre os participantes, não fez classificações da prática de atividades físicas e não definiu um PCU.

5.3.4 Identificação e compreensão das áreas desportivas urbanas

Ferrari e Mamei (2013) fizeram um estudo com a mineração dos dados de atividades físicas praticadas pelos usuários do Nokia Sports Tracker Service, disponibilizadas no repositório *online* da Nokia, onde foi feito o estudo das áreas urbanas onde houve a concentração da prática das atividades físicas (Ferrari e Mamei 2013). Assim, esse estudo gerou um mapeamento urbano da prática de atividades físicas, incluindo filtros de tipos de atividades físicas e o público que executou a atividade física.

A semelhança entre esse estudo e os arcabouços desenvolvidos nesta tese está no fato de ambos analisarem a prática de atividades físicas com o uso de AS de monitoramento de atividades físicas.

As diferenças são que o estudo não desenvolveu um aplicativo social para promover a prática de atividades físicas, não fez a coleta dessas informações no Facebook e baseou suas análises apenas em um único AS de monitoramento de atividades físicas.

5.4 Aplicativos Sociais de monitoramento de atividades físicas

Os Aplicativos Sociais de monitoramento de atividades físicas trabalham com o registro da prática das atividades físicas. Cada AS permite a criação de amizades entre os seus usuários, onde pode ser feito um *ranking* competitivo de corridas entre os mesmos. As atividades físicas registradas podem ser compartilhadas nos perfis dos usuários e permitem um maior nível de socialização e de interação do usuário.

As semelhanças entre os AS de monitoramento de atividades físicas e os arcabouços desenvolvidos nesta tese estão nos fatos de ambos estimularem a prática de atividades físicas com o uso de AS de monitoramento de atividades físicas e ambos compartilharem informações da prática de atividades físicas nas RSO.

As diferenças são que os AS de monitoramento de atividades físicas não fizeram a coleta das informações no Facebook, bem como não fizeram a definição do PCU.

5.5 Considerações Finais

Os trabalhos relacionados mostraram alguns estudos interessantes com base nas RSO. Alguns deles utilizaram a forma similar de coleta dos dados com o uso de Aplicativos Sociais, mas poucos utilizaram uma forma automática de coleta ou utilizaram o AS de Coleta dos Dados como uma ferramenta aplicada a algum interesse do usuário, como uma competição social ou a definição de um PCU.

Em relação às classificações, apesar de que alguns trabalhos utilizaram o processamento paralelo e distribuído, o foco da pesquisa foi diferente, buscando resultados diferentes.

Assim, não foi encontrado trabalho relacionado que faça exatamente o que foi proposto por essa tese, na forma de um *Framework* de análise do comportamento. Há soluções próximas, com o uso de gamificação para estimular a prática de atividades físicas, mas não continham sistemas de avaliação para detectar os excessos da prática da atividade física, o que é um comportamento pouco saudável e o seu praticante deveria ser alertado sobre os riscos inerentes a esse comportamento inadequado. A Tabela 6 sumariza os trabalhos relacionados e indica quais métricas cada um possui.

Tabela 6 - Resumo dos trabalhos relacionados

Trabalho	Métricas								
	Gamificação	PCU	Facebook	Competição Social	Aplicativo Social	Paralelismo	Análise de RSO	Foco em Fitness	Motivação esportiva
Nair et al. (2017)						X	X		
Vaterlaus et al. (2015)			X				X	X	X
Zang et al. (2015)			X				X	X	X
Cavallo et al. (2012)			X				X	X	X
Xue et al. (2010)			X			X	X		
Burke et al (2010)		X	X			X	X		
Stragier et al. (2016)		X			X		X	X	X
Foster et al. (2010)	X		X	X	X		X	X	X
MovaMais (2016)	X							X	X
Centola e Rijt (2015)	X							X	X
Lu et al. (2014)								X	X
Ferrari e Mamei (2013)								X	X
AS de monitoramento de atividades físicas	X				X			X	X
Arcabouços da tese	X	X	X	X	X	X	X	X	X

Capítulo 6 - Conclusões

Essa tese criou arcabouços computacionais que permitiram a coleta, a análise e a classificação dos dados dos usuários do Facebook, com o processamento paralelo e distribuído, para a definição do Padrão Comportamental do Usuário.

O contexto de classificação foi o da prática de atividades físicas, as quais foram compartilhadas nos perfis pelos Aplicativos Sociais de monitoramento de atividades físicas, para a definição do quão saudável está o comportamento do usuário.

Assim, o *Framework* foi desenvolvido com o contexto de análise de *fitness*. Foram modelados os *Workflows* para a coleta e para a análise dos dados. O arcabouço de Coleta dos Dados é um Aplicativo Social de acesso público para motivar à prática de atividades físicas, o qual fez a avaliação de seus usuários por 8 meses e considerou as informações da prática de atividades físicas feitas há até 4 anos.

Foi criado o modelo de análise por Panoramas, os quais descreveram Eventos associados aos Usuários em um determinado período de tempo, como a prática de atividades físicas em um determinado período.

Dado o pouco tempo de utilização do arcabouço de Coleta de Dados, em conjunto com uma quantidade reduzida de usuários, não foi possível coletar, em tempo hábil, um volume de dados que representasse um problema de *BigData*. Porém, a avaliação de desempenho sugeriu que a metodologia foi aplicada de maneira adequada para suportar o processamento paralelo e distribuído com o volume de dados de um *BigData*. Dessa forma, a arquitetura paralela e distribuída utilizada será adequada para as futuras demandas de aumento do volume do processamento dos dados, criando uma base de dados que permitirá a mineração de seus dados, o que possibilitará novas investigações sobre esses dados.

O foco motivacional de uso do arcabouço de Coleta dos Dados, voltado ao estímulo de um comportamento saudável com o uso de uma competição social saudável, torna lúdica ao usuário a mudança comportamental e, com o passar do tempo, formará uma base de dados com o volume de um *BigData*.

É possível fazer a análise de outros PCU, com focos diferentes de *fitness*, com ajustes nos arcabouços computacionais, incluindo o ajuste do modelo de classificação adequado ao novo PCU a ser analisado.

6.1 Contribuições e Publicações

As contribuições dessa tese estão relacionadas com a disponibilização do *Framework* para a classificação do comportamento saudável do usuário, de forma escalável com o aumento da demanda de utilização do mesmo.

O *Framework* de Coleta dos Dados está disponível para o uso público, com o viés motivacional de proporcionar uma competição social saudável entre seus usuários. Com essa contribuição é esperada uma maior socialização entre os seus usuários, o que poderá motivar a diminuição do estilo de vida sedentário nos usuários e em seus familiares, amigos e conhecidos.

Durante o desenvolvimento da tese, foram redigidos artigos relativos à proposta da tese e também aos artefatos intermediários produzidos, os quais tiveram contribuições significativas.

Até o presente momento, foram publicados 6 artigos em congressos internacionais, relatados a seguir na Tabela 7.

Tabela 7 – Publicações relacionadas com a tese

DOI: https://doi.org/10.23919/CISTI.2017.7975688 QUALIS: B1 (2015 - Ciência da Computação) PASCHOAL JÚNIOR, FABIO; EBECKEN, NELSON FRANCISCO FAVILLA; RIBEIRO, GABRIEL VINICIUS SILVA; DE ARAGAO DAQUER, LEANDRO MONIZ; MAURO, RENATO CAMPOS; OGASAWARA, EDUARDO SOARES. FitRank – Social App to Combat Physical Inactivity: Study of the use of Fitness Social Apps on Facebook’s users profiles. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), 2017, Lisboa. 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), 2017.
QUALIS: B1 (2007 - Engenharias I) PASCHOAL JÚNIOR, FÁBIO; RIBEIRO, GABRIEL VINICIUS SILVA; DAQUER, LEANDRO MONIZ DE ARAGÃO; MAURO, RENATO CAMPOS; OGASAWARA, EDUARDO SOARES; EBECKEN, NELSON FRANCISCO FAVILLA. IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS DE USUÁRIOS DE APLICATIVOS SOCIAIS PARA A PREDIÇÃO DO COMPORTAMENTO SAUDÁVEL DO USUÁRIO. In: XXXVII Iberian Latin American Congress on Computational Methods in Engineering, 2016, Brasília.
DOI: http://dx.doi.org/10.1109/CISTI.2016.7521484 QUALIS: B1 (2015 - Ciência da Computação) PASCHOAL, FABIO; EBECKEN, NELSON FRANCISCO FAVILLA; RIBEIRO, GABRIEL VINICIUS SILVA; DE ARAGAO DAQUER, LEANDRO MONIZ; MAURO, RENATO CAMPOS; OGASAWARA, EDUARDO SOARES. Healthy behavior with social apps: Proposal for evolution study of the use of fitness social apps on Facebook. In: 2016 11th Iberian Conference

on Information Systems and Technologies (CISTI), 2016, Gran Canaria. 2016 11th Iberian Conference on Information Systems and Technologies (CISTI), 2016.
DOI: http://dx.doi.org/10.20906/CPS/CILAMCE2015-0684 QUALIS: B1 (2007 - Engenharias I) PASCHOAL JÚNIOR, FÁBIO; RIBEIRO, GABRIEL VINICIUS SILVA; DAQUER, LEANDRO MONIZ DE ARAGÃO; MAURO, RENATO CAMPOS; OGASAWARA, EDUARDO SOARES; EBECKEN, NELSON FRANCISCO FAVILLA. FITRANK - DESENVOLVIMENTO DE APLICATIVO PARA A IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS DE ATIVIDADES FÍSICAS A PARTIR DE APLICATIVOS PARA REDES SOCIAIS. In: XXXVI Iberian Latin American Congress on Computational Methods in Engineering, 2015, Rio de Janeiro.
DOI: http://dx.doi.org/10.20906/CPS/CILAMCE2015-0771 QUALIS: B1 (2007 - Engenharias I) PASCHOAL JÚNIOR, FÁBIO; RIBEIRO, GABRIEL VINICIUS SILVA; DAQUER, LEANDRO MONIZ DE ARAGÃO; EBECKEN, NELSON FRANCISCO FAVILLA. PROPOSTA DE IDENTIFICAÇÃO DA EVOLUÇÃO TEMPORAL DE USO DE APLICATIVOS SOCIAIS PARA A DEFINIÇÃO DE PADRÕES COMPORTAMENTAIS DOS USUÁRIOS. In: XXXVI Iberian Latin American Congress on Computational Methods in Engineering, 2015, Rio de Janeiro.
QUALIS: B1 (2007 - Engenharias I) PASCHOAL JÚNIOR, F.; EBECKEN, N. F. F. . UMA ABORDAGEM PARA IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS A PARTIR DE APLICATIVOS PARA REDES SOCIAIS. In: CILAMCE2014 - XXXV Ibero Latin American Congress on Computational Methods in Engineering, 2014, Fortaleza. CILAMCE2014 - XXXV Ibero Latin American Congress on Computational Methods in Engineering, 2014.
PASCHOAL JÚNIOR, F.; MOREIRA NETO, P. C.; EBECKEN, N. F. F.. Contador de palavras paralelizado com o uso do MPI. In: 4º Congresso Científico da Semana Tecnológica IFSP, 2013, Bragança Paulista. Anais do Concistec'13. São Paulo: IFSP, 2013.

Há também um artigo em avaliação, enviado para a Revista Brasileira de Medicina do Esporte (Brazilian Journal of Sports Medicine) - Journal Impact Factor: 0.173 (JCR 2015 - Sport Sciences) - Eigenfactor Score: 0.000320 (JCR 2015 - Sport Sciences) - Qualis: A2 (2014 - Educação Física).

PASCHOAL JÚNIOR, FÁBIO; EBECKEN, NELSON FRANCISCO FAVILLA; RIBEIRO, GABRIEL VINICIUS SILVA; DE ARAGÃO DAQUER, LEANDRO MONIZ; MAURO, RENATO CAMPOS; OGASAWARA, EDUARDO SOARES. Nível de atividade física dos usuários do Facebook.

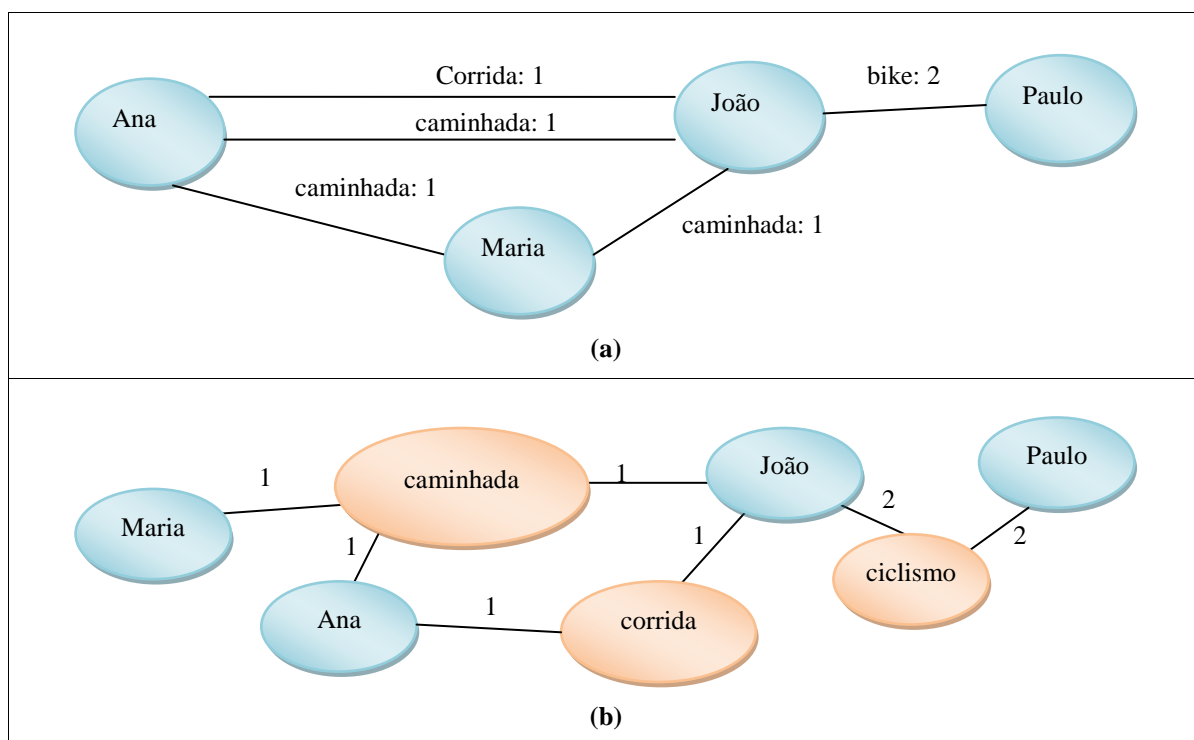
6.2 Trabalhos Futuros

Alguns trabalhos futuros são sugeridos, como continuação ou melhoria dos arcabouços computacionais desenvolvidos e das técnicas utilizadas. Um trabalho relativamente simples será disponibilizar o resultado dos PCU, de forma individual, ao usuário do arcabouço de Coleta dos Dados, através de ajustes da interface para disponibilizar somente ao mesmo, de forma privada, o seu PCU. A interface do FitRank será ajustada em um futuro próximo.

Dada a riqueza das informações armazenadas, é possível desenvolver alguns estudos complementares a esta tese. As próximas subseções ilustram os mesmos.

6.2.1 Visualizações de classificação dos usuários

É possível produzir diferentes visualizações com novos PCU, de acordo com o contexto desejado. Através das similaridades das execuções das atividades físicas, temos algumas variações de possibilidades ilustradas na Figura 51, dentre outras possibilidades, utilizando a técnica de subgrafos (Freitas et al. 2008b, Paschoal Jr et al. 2016a).



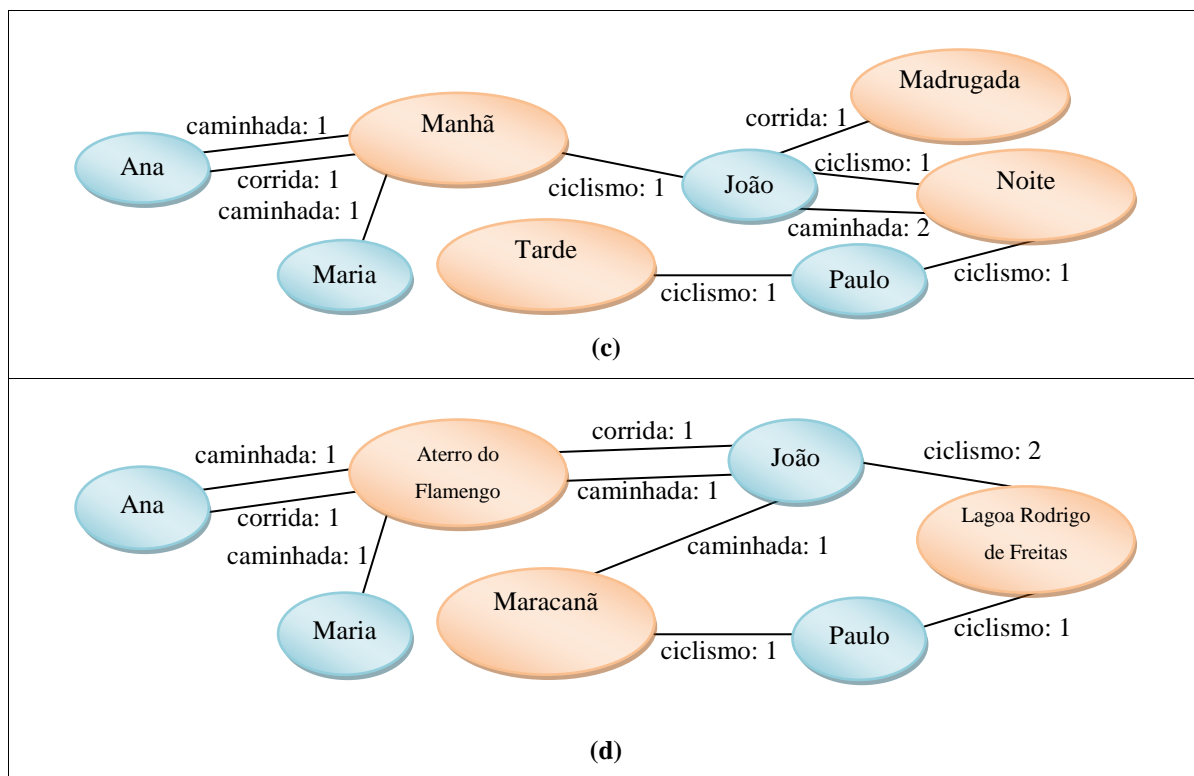


Figura 51 – Possíveis visualizações

A Figura 51(a) correlaciona os usuários com as atividades físicas e as suas quantidades executadas. Os usuários formam os vértices deste grafo. As atividades físicas e as suas quantidades de execução pelos usuários formam as arestas deste grafo.

A Figura 51(b) correlaciona os usuários com as atividades físicas e as suas quantidades executadas. As atividades físicas e os usuários formam os vértices deste grafo. As quantidades de execução das atividades físicas pelos usuários formam as arestas deste grafo (Paschoal Jr et al. 2016a).

A Figura 51(c) correlaciona os usuários com o período do dia de execução das atividades físicas. Os períodos do dia e os usuários formam os vértices deste grafo. As atividades físicas e suas respectivas quantidades praticadas formam as arestas deste grafo (Paschoal Jr et al. 2016a).

A Figura 51(d) correlaciona os usuários com os locais de realização das atividades físicas. Os locais e os usuários formam os vértices desse grafo. As atividades físicas e suas respectivas quantidades praticadas formam as arestas deste grafo (Paschoal Jr et al. 2016a).

6.2.2 Mineração dos Dados de Geo-referenciamento

De forma complementar, a mineração dos dados de geo-referenciamento permitirão novas descobertas e novas correlações de dados. Alguns possíveis exemplos, dentre outras possibilidades, são definir um público de esportistas de uma localidade em específico, ou a maior frequência de execução das atividades físicas (dias e horários), ou a distribuição das localidades de execução das atividades físicas, bem como o quanto uma localidade é utilizada para a prática de atividades físicas, ou até mesmo o estudo de séries temporais com a combinação de dados de geolocalização (Paschoal Jr et al. 2016a).

Haverá a necessidade de ver quais as técnicas de pré-processamento de dados serão mais adequadas ao trabalho. Com essas descobertas será possível planejar e executar ações públicas ou privadas, de forma que seja feito um melhor gerenciamento da prática de atividades físicas e de uma melhor utilização desses locais, inclusive com as necessidades de segurança pública ou de criar ações específicas para esse público alvo desejado (Paschoal Jr et al. 2016a).

6.2.3 Predição do Comportamento do Usuário

A predição do comportamento saudável do usuário poderá ser feita de duas formas. Em um primeiro momento, indicará a situação atual do usuário. Em um segundo momento, indicará o futuro comportamento do usuário se ele mantiver as atuais frequência e intensidade nas suas atividades físicas (Paschoal Jr et al. 2016a).

Dessa forma, a predição definirá o atributo de comportamento saudável do usuário com os valores “pouco saudável”, “saudável” e “muito saudável”, além de que poderá recomendar um novo comportamento para o usuário no sentido de melhorar a sua situação, onde recomendará no que poderá ser melhorado em suas atividades físicas, como frequência, duração, quantidade ou desempenho (Paschoal Jr et al. 2016a).

6.2.4 Evolução temporal de histórico do Usuário

As informações dos PCU poderão ser tratadas na forma de uma evolução temporal do usuário, onde será possível a criação de um histórico da evolução de execução de

atividades físicas (Paschoal Jr et al. 2016a, 2016b).

Esse histórico poderá ser utilizado para distintas finalidades. Algumas possibilidades de utilização são: i.) a obtenção, para o usuário, de algum tipo de desconto em planos privados de saúde com base no perfil traçado de prevenção de doenças; ii.) um melhor acompanhamento médico da saúde do usuário e; iii.) a criação de um plano de recompensas de acordo com a quantidade e a qualidade das atividades físicas executadas (Paschoal Jr et al. 2016a, 2016b).

Além disso, poderá ser estudada a influência temporal de quanto um amigo estimulou o outro amigo à prática de atividades físicas.

6.2.5 Uso de *Machine Learning*

Os algoritmos de *Machine Learning* poderão ser utilizados nas classificações, nos agrupamentos e nas sumarizações dos dados. A classificação poderá utilizar o algoritmo de classificação *Naive Bayes*, com o paradigma de *MapReduce*.

O agrupamento e a sumarização poderão utilizar o algoritmo de clusterização *K-Means Clustering*, com o paradigma de *MapReduce*. Linguagens como R e Python poderão ser utilizadas em conjunto com o Apache Spark e com os algoritmos de *Machine Learning*, para proporcionar novas análises dos dados, pelo motivo dessas linguagens oferecem amplas bibliotecas de *Machine Learning*.

6.2.6 Coleta de dados complementares com Jsoup

A maioria dos AS de atividades físicas postam os seus conteúdos nos perfis dos usuários do Facebook de forma bem resumida, porém contendo um hipertexto para o *site* externo do AS, o qual detalha melhor a execução das atividades físicas. Como essas informações detalhadas não estão no Facebook, elas poderão ser coletadas com o uso de outra API não pertencente ao Facebook. Então, a API Jsoup possibilita a coleta de dados de páginas da *web*, é uma biblioteca feita em Java e pode ser utilizada para coletar essas informações detalhadas disponibilizadas no hipertexto para o *site* externo do AS (Hedley 2017, Paschoal Jr et al. 2016a).

Essa coleta adicional com o Jsoup enriquecerá os dados das atividades físicas já coletadas pelo arcabouço de Coleta dos Dados, possibilitando novas minerações de dados com novas descobertas de informações.

Referências bibliográficas

- Allen, M., Bartels, N., (2017). RestFB. Disponível em: <http://restfb.com/>. Acesso em: 21 maio 2017.
- Alves, G., Antunes, J., (2015), "Novo Paradigma na Comunicação - As Redes Sociais entre Marcas e Consumidores", *Atas da 10ª Conferência Ibérica de Sistemas e Tecnologias de Informação – CISTI 2015*, v. Vol. I – Artigos (jun.), p. 769–773.
- Apache Spark, (2017a). Apache Spark™ - Lightning-Fast Cluster Computing. Disponível em: <http://spark.apache.org/>. Acesso em: 22 maio 2017.
- Apache Spark, (2017b). Spark Programming Guide - Spark 2.1.1 Documentation. Disponível em: <http://spark.apache.org/docs/latest/programming-guide.html>. Acesso em: 22 maio 2017.
- Apache Spark, (2017c). Spark SQL and DataFrames - Spark 2.1.1 Documentation. Disponível em: <http://spark.apache.org/docs/latest/sql-programming-guide.html#datasets-and-dataframes>. Acesso em: 23 maio 2017.
- Apache Spark, (2017d). Dataset - org.apache.spark.sql.Dataset. Disponível em: <http://spark.apache.org/docs/latest/api/scala/index.html#org.apache.spark.sql.Dataset>. Acesso em: 23 maio 2017.
- Apache Spark, (2017e). Spark SQL & DataFrames | Apache Spark. Disponível em: <http://spark.apache.org/sql/>. Acesso em: 22 maio 2017.
- Aquino, F., Gomes, M., (2016). Mova Mais. Disponível em: <http://www.movamais.com>. Acesso em: 21 maio 2017.
- Barbosa, G. A. R., Santos, G. E. dos, Pereira, V. M. de O., (2013), "Caracterização qualitativa da sociabilidade no Facebook", *IHC '13 Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems*, p. 72–81.
- Benevenuto, F., (2010), "Redes Sociais Online: Técnicas de Coleta, Abordagens de Medição e Desafios Futuros", *Short course (book chapter) in SBSC'10, Webmedia'10, IHC'10 and SBBD'10*
- Benevenuto, F., Almeida, J. M., Silva, A. S., (2012), "Coleta e Análise de Grandes Bases de Dados de Redes Sociais Online", *Book Chapter in Jornada de Atualizações em Informática (JAI), Congresso da Sociedade Brasileira de Computação (CSBC)*

- Benevenuto, F., Almeida, V., (2011), "Uma Análise Empírica de Interações em Redes Sociais", *Proceedings of the XXIV Concurso de teses e dissertações (CTD)*
- boyd, danah m., Ellison, N. B., (2008), "Social network sites: Definition, history, and scholarship", *Journal of Computer-Mediated Communication*, v. 13, p. 210–230.
- Burke, M., Marlow, C., Lento, T., (2010), "Social Network Activity and Social Well-Being", *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, p. 1909–1912.
- Cavallo, D. N., Tate, D. F., Ries, A. V., Brown, J. D., DeVellis, R. F., Ammerman, A. S., (2012), "A Social Media-Based Physical Activity Intervention: A Randomized Controlled Trial", *American Journal of Preventive Medicine*, v. 5, n. 43, p. 527–532.
- Centola, D., Rijt, A. van de, (2015), "Choosing your network: Social preferences in an online health community", *Social Science & Medicine*, n. 125, p. 19–31.
- Cervi, C. R., (2008), "Um Estudo sobre Mineração de Dados em Redes Sociais", *Universidade Federal do Rio Grande do Sul*
- Chapman, C. N., Lahav, M., (2008), "International ethnographic observation of social networking sites", *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, p. 3123–3128.
- Chau, D. H., Pandit, S., Wang, S., Faloutsos, C., (2007), "Parallel crawling for online social networks", *World Wide Web Conference (WWW)*, p. 1283–1284.
- Cheng, Y., Park, J., Sandhu, R., (2013), "Preserving user privacy from third-party applications in online social networks", *Proceedings of the 22nd international conference on World Wide Web companion (WWW' 13)*
- Coelho-Ravagnani, C. de F., Melo, F. C. L., Ravagnani, F. C. P., Burini, F. H. P., Burini, R. C., (2013), "Estimativa do equivalente metabólico (MET) de um protocolo de exercícios físicos baseada na calorimetria indireta", *Rev Bras Med Esporte*, v. 19, n. 2 (mar.), p. 134–138.
- comScore, (2012a). Facebook Blasts into Top Position in Brazilian Social Networking Market Following Year of Tremendous Growth. *comScore, Inc.* Disponível em: http://www.comscore.com/Insights/Press_Releases/2012/1/Facebook_Blasts_into_Top_Position_in_Brazilian_Social_Networking_Market. Acesso em: 27 jun 2016.
- comScore, (2012b). Argentina Ranks First in Worldwide Desktop Social Networking Engagement at Nearly 10 Hours per Visitor Each Month. *comScore, Inc.*

- Disponível em: http://www.comscore.com/por/Insights/Press_Releases/2012/12/Argentina_Ranks_First_in_Worldwide_Desktop_Social_Networking_Engagem. Acesso em: 27 jun 2016.
- Consolvo, S., Everitt, K., Smith, I., Landay, J. A., (2006), "Design requirements for technologies that encourage physical activity", *CHI 2006 Proceedings - Designing for Tangible Interactions*, p. 457–466.
- Cosenza, V., (2017). World Map of Social Networks. Disponível em: <http://vincos.it/world-map-of-social-networks/>. Acesso em: 21 maio 2017.
- Dean, J., Ghemawat, S., (2008), "MapReduce: Simplified Data Processing on Large Clusters", *Commun. ACM*, v. v. 51, n. n. 1 (jan.), p. 107–113.
- Domingues, F. L., (2008). Computação ubíqua. Disponível em: <http://www.hardware.com.br/artigos/computacao-ubiqua/>. Acesso em: 15 fev 2017.
- Ellison, N. B., Steinfield, C., Lampe, C., (2007), "The benefits of Facebook “friends”: Exploring the relationship between college students’ use of online social networks and social capital", *Journal of Computer-Mediated Communication*, v. 12, p. 1143–1168.
- Endomondo, (2017a). Endomondo Sports Tracker. Disponível em: <https://www.facebook.com/games/endoapp/?fbs=133>. Acesso em: 21 maio 2017.
- Endomondo, (2017b). Endomondo Corrida Ciclismo MTB. Disponível em: <https://play.google.com/store/apps/details?id=com.endomondo.android>. Acesso em: 21 maio 2017.
- Facebook, (2014a). Central de Aplicativos. Disponível em: <https://www.facebook.com/appcenter/>. Acesso em: 1 jul 2014.
- Facebook, (2014b). Keyword Insights API. Disponível em: https://developers.facebook.com/docs/keyword_insights/. Acesso em: 1 jul 2014.
- Facebook, (2014c). Aplicativos: Saúde e Fitness. Disponível em: <https://www.facebook.com/appcenter/category/healthfitness/?platform=allplatforms>. Acesso em: 1 jul 2014.
- Facebook, (2014d). Seus aplicativos. Disponível em: <https://www.facebook.com/appcenter/my>. Acesso em: 1 jul 2014.
- Facebook, (2016a). Creating Custom Open Graph Stories. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/sharing/opengraph/custom>. Acesso em: 27 jun 2016.

- Facebook, (2016b). Chat API (Deprecated). *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/chat/>. Acesso em: 26 out 2016.
- Facebook, (2016c). Facebook Query Language (FQL) Reference. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/reference/fql>. Acesso em: 26 out 2016.
- Facebook, (2017a). The Graph API. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/graph-api>. Acesso em: 21 maio 2017.
- Facebook, (2017b). Plug-ins Sociais. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/plugins>. Acesso em: 21 maio 2017.
- Facebook, (2017c). Compartilhamento. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/sharing>. Acesso em: 21 maio 2017.
- Facebook, (2017d). Log de alterações da Plataforma do Facebook. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/apps/changelog>. Acesso em: 21 maio 2017.
- Facebook, (2017e). Facebook Platform Versioning. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/apps/versions>. Acesso em: 21 maio 2017.
- Facebook, (2017f). Outras APIs. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/graph-api/other-apis/>. Acesso em: 21 maio 2017.
- Facebook, (2017g). API de Feed Público. *Facebook Developers*. Disponível em: https://developers.facebook.com/docs/public_feed/. Acesso em: 21 maio 2017.
- Farinatti, P. de T. V., (2003), "Apresentação de uma Versão em Português do Compêndio de Atividades Físicas: uma contribuição aos pesquisadores e profissionais em Fisiologia do Exercício", *Revista Brasileira de Fisiologia do Exercício*, v. 2, p. 177–208.
- Fedkiw, R., (2012). Sensors and Cellphones. Disponível em: <http://www.stanford.edu/class/cs75n/Sensors.pdf>. Acesso em: 26 out 2016.
- Ferrari, L., Mamei, M., (2013), "Identifying and understanding urban sport areas using Nokia Sports Tracker", *Pervasive and Mobile Computing*, v. 9, p. 616–628.
- FitRank, (2017a). FitRank. Disponível em: <http://eic.cefet-rj.br/app/FitRank/>. Acesso em: 21 maio 2017.
- FitRank, (2017b). FitRank. Disponível em: <https://www.facebook.com/fitrank.go/>. Acesso em: 21 maio 2017.

- Foster, D., Linehan, C., Kirman, B., Lawson, S., James, G., (2010), "Motivating physical activity at work: Using persuasive social media for competitive step counting", *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, p. 111–116.
- Freitas, C. M. D. S., Nedel, L. P., Galante, R., Lamb, L. C., Spritzer, A. S., Fujii, S., Oliveira, J. P. M. de, Araújo, R. M., Moro, M. M., (2008a), "Extração de Conhecimento e Análise Visual de Redes Sociais", *Anais do XXVIII Congresso da SBC – SEMISH – Seminário Integrado de Software e Hardware*, p. 106–120.
- Freitas, C. M. D. S., Nedel, L. P., Galante, R., Lamb, L. C., Spritzer, A. S., Fujii, S., Oliveira, J. P. M. de, Araújo, R. M., Moro, M. M., (2008b), "Extração de Conhecimento e Análise Visual de Redes Sociais", *Anais do XXVIII Congresso da SBC – SEMISH – Seminário Integrado de Software e Hardware*, p. 106–120.
- Google, (2017). Google Play Store. Disponível em: <https://play.google.com/store>. Acesso em: 21 maio 2017.
- Han, J., Pei, J., Kamber, M., (2011), *Data Mining: Concepts and Techniques, Third Edition*. 3 edition ed. MA, Morgan Kaufmann.
- Hedley, J., (2017). jsoup Java HTML Parser, with best of DOM, CSS, and jquery. Disponível em: <http://jsoup.org/>. Acesso em: 21 maio 2017.
- Hurling, R., Catt, M., Boni, M. D., Fairley, B. W., Hurst, T., Murray, P., Richardson, A., Sodhi, J. S., (2007), "Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial", *Journal of Medical Internet Research*, v. 9(2), n. e7
- Hypeness, (2012). Case “Corrida Nike SP-Rio” foi Bronze no Facebook Studio Award 2012. Disponível em: <http://www.hypeness.com.br/2012/05/corrida-nike-sp-rio/>. Acesso em: 26 out 2016.
- Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., Shahabi, C., (2014), "Big data and its technical challenges", *Communications of the ACM*, v. v. 57, n. n. 7, p. 86–94.
- Joinson, A. N., (2008), "Looking at, looking up or keeping up with people?: motives and use of facebook", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1027–1036.
- Karau, H., Konwinski, A., Wendell, P., Zaharia, M., (2015), *Learning Spark - Lightning-fast Data Analysis*. O'Reilly Media.

- Kraft, P., Drozd, F., Olsen, E., (2008), "Digital Therapy: Addressing Willpower as Part of the Cognitive-Affective Processing System in the Service of Habit Change", *PERSUASIVE 2008*, p. 177–188.
- Kramer, A. D. I., (2012), "The spread of emotion via facebook", *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, p. 767–770.
- Lampe, C., Ellison, N. B., Steinfield, C., (2008), "Changes in use and perception of facebook", *Proc. Of the 2008 ACM conference on Computer supported cooperative work (CSCW '08)*, p. 721–730.
- Lu, Y., Kim, Y., Dou, X. (Yuki), Kumar, S., (2014), "Promote physical activity among college students: Using media richness and interactivity in web design", *Computers in Human Behavior*, v. 41, p. 40–50.
- Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., Beaton, M., (2013), "Teens, Social Media, and Privacy", *PewResearchCenter – The Berkman Center for Internet & Society at Harvard University*
- Maksimovic, M., Vujovic, V., Perisic, B., (2015), "A Custom Internet of Things Healthcare System", *Atas da 10^a Conferência Ibérica de Sistemas e Tecnologias de Informação – CISTI 2015*, v. Vol. I – Artigos (jun.), p. 653–658.
- MapMyFitness, (2017a). MapMyFitness. Disponível em: <https://www.facebook.com/mapmyfitness/?fref=ts>. Acesso em: 21 maio 2017.
- MapMyFitness, (2017b). Map My Fitness – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=com.mapmyfitness.android2>. Acesso em: 21 maio 2017.
- MapMyRide, (2017a). MapMyRide. Disponível em: <https://www.facebook.com/mapmyride/?fref=ts>. Acesso em: 21 maio 2017.
- MapMyRide, (2017b). MapMyRide Equitação Ciclismo – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=com.mapmyride.android2>. Acesso em: 21 maio 2017.
- MapMyRun, (2017a). MapMyRun. Disponível em: <https://www.facebook.com/mapmyrun/?fref=ts>. Acesso em: 21 maio 2017.
- MapMyRun, (2017b). Corra com MapMyRun – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=com.mapmyrun.android2>. Acesso em: 21 maio 2017.
- MapMyWalk, (2017a). MapMyWalk. Disponível em: <https://www.facebook.com/mapmywalk/?fref=ts>. Acesso em: 21 maio 2017.

- MapMyWalk, (2017b). Map My Walk – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=com.mapmywalk.android2>. Acesso em: 21 maio 2017.
- Mislove, A. E., (2009), "Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems", *Rice University*
- Morris, M. E., Consolvo, S., Munson, S., Patrick, K., Tsai, J., Kramer, A. D. I., (2011), "Facebook for health: opportunities and challenges for driving behavior change", *Proceeding CHI EA '11 CHI '11 Extended Abstracts on Human Factors in Computing Systems (nov.)*, p. 443–446.
- Munson, S. A., Lauterbach, D., Newman, M. W., Resnick, P., (2010), "Happier Together: Integrating a Wellness Application Into a Social Network Site", *Proc. of Persuasive 2010*
- Mutawa, N. A., Baggili, I., Marrington, A., (2012), "Forensic analysis of social networking applications on mobile devices", *Digital Investigation*, v. 9 (ago.), p. S24–S33.
- Nair, L. R., Shetty, S. D., Shetty, S. D., (2017), "Applying spark based machine learning model on streaming big data for health status prediction", *Computers and Electrical Engineering*
- Nazir, A., Raza, S., Chuah, C.-N., (2008), "Unveiling facebook: a measurement study of social network based applications", *ACM SIGCOMM Conference on Internet Measurement (IMC)*, p. 43–56.
- Newman, M. W., Lauterbach, D., Munson, S. A., Resnick, P., Morris, M. E., (2011), "It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and Opportunities in Using Online Social Networks for Health", *CSCW 2011*, p. 341–350.
- Nike, (2017a). Nike+ Running - get more from your run. Disponível em: <https://www.facebook.com/games/nikeapp/?fbs=133>. Acesso em: 21 maio 2017.
- Nike, (2017b). Nike+ Running. Disponível em: https://play.google.com/store/apps/details?id=com.nike.plusgps&hl=pt_BR. Acesso em: 21 maio 2017.
- Ogasawara, E. S., (2011), *UMA ABORDAGEM ALGÉBRICA PARA WORKFLOWS CIENTÍFICOS COM DADOS EM LARGA ESCALA*. Tese de Doutorado, Universidade Federal do Rio de Janeiro - COPPE
- Oliveira, T. H. M. de, Painho, M., (2015), "Emotion & Stress Mapping: Assembling an Ambient Geographic Information-based methodology in order to understand

- Smart Cities", *Atas da 10ª Conferência Ibérica de Sistemas e Tecnologias de Informação – CISTI 2015*, v. Vol. II – Artigos Curtos, Artigos Poster, Simpósio Doutoral (jun.), p. 351–354.
- Olsen, E., Kraft, P., (2009), "ePsychology: A pilot study on how to enhance social support and adherence in digital interventions by characteristics from social networking sites", *Persuasive 2009*
- Pantas, W., (2013). Saiba mais sobre os sensores em smartphones e tablets. Disponível em: <http://understech.com.br/saiba-mais-sobre-os-sensores-em-smartphones-e-tablets/>. Acesso em: 26 out 2016.
- Paschoal Jr, F., Ebecken, N. F. F., (2014), "Uma abordagem para identificação de padrões comportamentais a partir de aplicativos para redes sociais", *Proceedings of the XXXV Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2014*
- Paschoal Jr, F., Ribeiro, G. V. S., Daquer, L. M. de A., Ebecken, N. F. F., (2015a), "Proposta de identificação da evolução temporal de uso de aplicativos sociais para a definição de padrões comportamentais dos usuários", *Proceedings of the XXXVI Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2015*
- Paschoal Jr, F., Ribeiro, G. V. S., Daquer, L. M. de A., Mauro, R. C., Ogasawara, E. S., Ebecken, N. F. F., (2015b), "FitRank - Desenvolvimento de aplicativo para a identificação de padrões comportamentais de atividades físicas a partir de aplicativos para redes sociais", *Proceedings of the XXXVI Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2015*
- Paschoal Jr, F., Ribeiro, G. V. S., Daquer, L. M. de A., Mauro, R. C., Ogasawara, E. S., Ebecken, N. F. F., (2016a), "Identificação de padrões comportamentais de usuários de aplicativos sociais para a predição do comportamento saudável do usuário", *Proceedings of the XXXVII Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2016*
- Paschoal Jr, F., Ribeiro, G. V. S., Daquer, L. M. de A., Mauro, R. C., Ogasawara, E. S., Ebecken, N. F. F., (2016b), "Comportamento Saudável com Aplicativos Sociais: Proposta de estudo de evolução do uso de aplicativos sociais de fitness no Facebook", *Actas de la 11a Conferencia Ibérica de Sistemas y Tecnologías de Información*, v. Vol. I – Artículos de la Conferencia (jun.), p. 917–922.

- Péres, F. de A., Rodríguez, M. C., Gago, J. M. S., (2015), "Knowledge extraction from usage data of mobile devices with educational purposes", *Atas da 10ª Conferência Ibérica de Sistemas e Tecnologias de Informação – CISTI 2015*, v. Vol. II – Artigos Curtos, Artigos Poster, Simpósio Doutoral (jun.), p. 326–329.
- Polar Beat, (2017a). Polar Beat. Disponível em: <https://www.facebook.com/games/polarbeatdomain/?fbs=-1>. Acesso em: 21 maio 2017.
- Polar Beat, (2017b). Polar Beat – Fitness Coach – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=fi.polar.beat>. Acesso em: 21 maio 2017.
- Polar Flow, (2017a). Polar Flow. Disponível em: https://flow.polar.com/?ref=br_rs. Acesso em: 21 maio 2017.
- Polar Flow, (2017b). Polar Flow - Activity & Sports – Apps para Android no Google Play. Disponível em: <https://play.google.com/store/apps/details?id=fi.polar.polarflow>. Acesso em: 21 maio 2017.
- RunKeeper, (2017a). RunKeeper. Disponível em: <https://www.facebook.com/games/runkeeper-og/?fbs=133>. Acesso em: 21 maio 2017.
- RunKeeper, (2017b). Runkeeper GPS Correr Caminhar. Disponível em: <https://play.google.com/store/apps/details?id=com.fitnesskeeper.runkeeper.pro>. Acesso em: 21 maio 2017.
- Runtastic, (2017a). Runtastic.com. Disponível em: <https://www.facebook.com/games/runtastic/?fbs=133>. Acesso em: 21 maio 2017.
- Runtastic, (2017b). Runtastic Corrida e Caminhada. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android&hl=pt_BR. Acesso em: 21 maio 2017.
- Runtastic, (2017c). Runtastic PRO Correr, Caminhar. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android.pro2&hl=pt_BR. Acesso em: 21 maio 2017.
- Runtastic, (2017d). Runtastic Road Bike Ciclismo. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.lite&hl=pt_BR. Acesso em: 21 maio 2017.
- Runtastic, (2017e). Runtastic Road Bike PRO GPS. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.pro&hl=pt_BR. Acesso em: 21 maio 2017.

- Runtastic, (2017f). Runtastic Mountain Bike GPS. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.lite&hl=pt_BR. Acesso em: 21 maio 2017.
- Runtastic, (2017g). Runtastic Mountain Bike PRO. Disponível em: https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.pro&hl=pt_BR. Acesso em: 21 maio 2017.
- Runtastic Shop, (2016). Runtastic Receiver and Heart Rate Monitor. Disponível em: <https://www.runtastic.com/shop/en/runtastic-receiver-heart-rate-monitor>. Acesso em: 26 out 2016.
- Runtastic Shop, (2017). Runtastic Heart Rate Combo Monitor. Disponível em: <https://www.runtastic.com/shop/en/runtastic-bluetooth-smart-combo-heart-rate-monitor>. Acesso em: 21 maio 2017.
- Schirmer, M., Höpfner, H., (2011). Smartphone Hardware Sensors. Disponível em: <https://www.uni-weimar.de/medien/wiki/images/Zeitmaschinen-smartphonesensors.pdf>. Acesso em: 26 out 2016.
- Silva, A. A. S. da, (2015), *NET-Y: Uma abordagem para detecção de atividades suspeitas em redes sociais*, Universidade Federal do Rio de Janeiro - COPPE Disponível em: http://www.coc.ufrj.br/index.php?option=com_docman&view=document&alias=2709-silva-aas-tm-15-pdf&category_slug=2015-1&Itemid=428.
- Silva, C., Hanne, M., Almeida, J., Gonçalves, M., (2012), "Analysis of vulnerability to facebook users", *In Proc. of the 18th Brazilian symposium on Multimedia and the web (WebMedia '12)*, p. 335–342.
- Stragier, J., Abeele, M. V., Mechant, P., Lieven De Marez, (2016), "Understanding persistence in the use of Online Fitness Communities: Comparing novice and experienced users", *Computers in Human Behavior*, v. 64, p. 34–42.
- Strava, (2017a). Strava. Disponível em: <https://www.facebook.com/games/stravainc/?fbs=-1>. Acesso em: 21 maio 2017.
- Strava, (2017b). Strava GPS Correr Ciclismo. Disponível em: <https://play.google.com/store/apps/details?id=com.strava>. Acesso em: 21 maio 2017.
- The Apache Software Foundation, (2017). Welcome to Apache™ Hadoop®! Disponível em: <http://hadoop.apache.org/>. Acesso em: 15 fev 2017.

- Vaterlaus, J. M., Patten, E. V., Roche, C., Young, J. A., (2015), "#Gettinghealthy: The perceived influence of social media on young adult health behaviors", *Computers in Human Behavior*, v. 45, p. 151–157.
- Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., Cranor, L. F., (2011), "'I regretted the minute I pressed share': a qualitative study of regrets on Facebook", *Proc. of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*
- Wikipedia, (2017), *List of social networking websites* Disponível em: https://en.wikipedia.org/wiki/List_of_social_networking_websites.
- World Health Organization, (2002). World Health Day | 2002: move for health. *WHO*. Disponível em: <http://www.who.int/world-health-day/previous/2002/en/>. Acesso em: 27 jun 2016.
- World Health Organization, (2010), *Global recommendations on physical activity for health*.
- World Health Organization, (2017). WHO | Physical activity. *WHO*. Disponível em: <http://www.who.int/mediacentre/factsheets/fs385/en/>. Acesso em: 15 abr 2017.
- Xue, W., Shi, J., Yang, B., (2010), "X-RIME: Cloud-Based Large Scale Social Network Analysis", *Proceedings 2010 IEEE International Conference on Services Computing*, p. 506–513.
- Zhang, J., Brackbill, D., Yang, S., Centola, D., (2015), "Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial", *Preventive Medicine Reports*, v. 2, p. 651–657.